华章数学译丛

23

HZ BOOKS

WILEY

Analysis of Financial Time Series

金融时间序列分析

(美) Ruey S. Tsay 著
芝加哥大学

潘家柱 译



机械工业出版社
China Machine Press

金融时间序列分析

本书综合介绍金融计量模型及其在金融时间序列数据建模和预测中的应用，可帮助读者了解金融数据的基本特征，掌握金融计量模型的应用，并获得分析金融时间序列的经验。

对计量经济学和统计学文献中金融计量方法方面的最新进展进行概述是本书的突出特点。这些进展包括当前的研究热点，如风险值、高频数据分析和马尔可夫链蒙特卡罗方法等。特别是一些在学术杂志上尚未发表的最新成果，如使用跳跃扩散方程进行衍生产品的定价，基于非齐次二维泊松过程的极值理论计算风险值，以及带时变相关系数的多元波动率模型。此外，本书还介绍了用MCMC方法进行金融中的贝叶斯推断。

强调实例和数据分析是本书的另一个突出特点。全书采用实际金融数据说明所讨论模型和方法的应用。建立线性时间序列模型用SCA；估计波动率模型用RATS（时间序列的回归分析）；实现神经网络和绘制附图用S-Plus。用Fortran程序进行简单的期权定价、估计极值模型、计算风险值和进行贝叶斯分析。

作 者 简 介

Ruey S. Tsay 于美国威斯康星大学麦迪逊分校获得统计学博士学位，美国芝加哥大学商学院研究生院经济计量及统计学的H.G.B. Alexander教授。曾任*Journal of Financial Econometrics* 杂志栏目编辑。



Analysis of Financial Time Series



www.wiley.com

华章网站 http://www.hzbook.com

网上购书：www.china-pub.com

投稿热线：(010) 88379604

购书热线：(010) 68995259, 68995264

读者信箱：hzjsj@hzbook.com





华章图书



ISBN 7-111-18386-X/O · 478

定价：39.00 元

封面设计 · 杨宇梅

本书主要介绍了在计量经济学和统计学文献中出现的金融计量方法方面的最新进展，强调实例和数据分析．特别是包含当前的研究热点，如风险值、高频数据分析和马尔可夫链蒙特卡罗方法等．主要内容包括：金融时间序列数据的基本特征，神经网络，非线性方法，使用跳跃扩散方程进行衍生产品的定价，采用极值理论计算风险值，带时变相关系数的多元波动率模型，贝叶斯推断．

本书可作为金融等专业高年级本科生或研究生的时间序列分析教材，也可供相关专业研究人员参考．

图书在版编目（CIP）数据

金融时间序列分析/（美）蔡（Tsay，R. S.）著；潘家柱译．-北京：机械工业出版社，2006.4
（华章数学译丛）
书名原文：Analysis of Financial Time Series
ISBN 7-111-18386-X

Ⅰ．金…　Ⅱ．①蔡…　②潘…　Ⅲ．时间序列分析－应用－金融－分析　Ⅳ．F83

中国版本图书馆 CIP 数据核字（2006）第 003941 号

机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码　100037）
责任编辑：王　玉
北京诚信伟业印刷有限公司印刷・新华书店北京发行所发行
2006 年 4 月第 1 版第 1 次印刷
787mm×1020mm 1/16・22.75 印张
定价：39.00 元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换
本社购书热线：（010）68326294

译 者 序

时间序列分析在理论和经验上已成为金融市场研究的不可缺少的部分. 时间序列分析方法已是金融定量分析的主流方法之一. 近代计量经济和金融市场的许多研究成果都建立在时间序列分析的基础之上. Engle 和 Grange 因为他们的时间序列模型在经济金融中的广泛应用而获得 2003 年的诺贝尔经济学奖，就是时间序列分析方法的重要性在世界上被广泛认可的有力证明.

蔡瑞胸(Ruey S. Tsay)教授是美国芝加哥大学的计量经济与统计学的 H. G. B. 亚历山大(Alexander)教授. 他在计量经济学、统计学和金融市场的研究方面成果卓著. 他的这本《金融时间序列分析》涵盖了当前数理金融研究中最新的几个重要方面：风险值的计算、用显式表示的跳跃扩散过程进行期权定阶、高频金融数据的分析、带时变相关系数的多元波动率模型、马尔可夫链蒙特卡罗(MCMC)方法等. 本书从金融时间序列的基本特征出发，讨论一元金融时间序列的分析和应用、多个资产的收益率以及金融模型的贝叶斯推断等问题. 全书都在利用实例和实际的金融数据，来说明所述模型和方法的应用，并附有计算程序. 这是一本比较全面系统地介绍金融计量模型，及其在金融时间序列数据的建模和预测中的应用的好书，在国外已被广泛引用.

译者几年来在北京大学金融数学系开设"金融时间序列分析"的课程，深知在课程中既做到理论和实际相结合又做到深入浅出的难度. 蔡教授在本书中做到了理论与实际相结合，在叙述方式上深入浅出，而且结合了许多最新的金融计量学的研究成果和方法. 本书对学习金融经济学、统计学的本科生和研究生来说，是一本很好的教科书；对管理、金融方面的业内人士和研究人员是一本很好的参考书. 我相信本书中译本的出版对我国的金融计量学和统计学的教学、对现代金融计量方法在我国金融界的应用、对我国金融市场的量化研究工作都是有益的.

刘淑霞协助翻译了第 5～10 章，王辉、郭沁苗、肖业英和李宛朋也参与了部分翻译工作. 没有他们的帮助本书的翻译工作是不可能完成的. 对北京大学金融数学系的同事们的理解和支持，在此一并表示衷心的感谢. 由于时间仓促，本人的精力和水平有限，翻译的谬误之处一定不少，望同行们和读者们多多指正.

潘家柱

2005 年 12 月

前　言

本书由自 1999 年我在芝加哥大学商学院所教的 MBA(工商管理硕士)金融时间序列分析课程发展而来. 它也包含了过去几年我开设的时间序列分析博士生课程的内容. 这是一本引论性质的书，旨在对金融计量模型及其在金融时间序列数据建模和预测中的应用，进行系统的、综合的阐述. 目标是使读者了解金融数据的基本特征，懂得金融计量模型的应用，并获得分析金融时间序列的经验.

本书可作为金融专业 MBA 学生的时间序列分析教材，也适用于商学、经济学、数学和统计学专业对金融计量学感兴趣的研究生和高年级本科生. 它也可以作为要进行风险值(Value at Risk)的计算、波动率(Volatility)建模和对具有先后相关性的数据进行分析等工作的研究人员和业内人士的参考书.

对计量经济学和统计学文献中的金融计量方法方面的最新进展进行概述是本书的突出特点. 这些进展包括当前的研究热点，如风险值(VaR)、高频数据分析和马尔可夫链蒙特卡罗(MCMC)方法等. 特别是，包含一些在学术杂志上尚未发表的最新成果，请参阅第 6 章中关于使用有显式公式的跳跃扩散方程来进行衍生产品的定价，第 7 章中基于非齐次二维泊松(Poisson)过程的极值理论计算风险值(VaR)，以及第 9 章中带时变相关系数的多元波动率模型. 之所以介绍 MCMC 方法，是因为这类方法在金融计量中很有威力，并有大量的应用，而且将来的应用范围会更加广泛.

强调实例和数据分析是本书的另一个突出特点. 全书采用实际金融数据来说明所讨论模型和方法的应用. 使用多种计算机软件进行分析；建立线性时间序列模型用 SCA(Scientific Computing Associate，科学计算助手)；估计波动率模型用时间序列的回归分析(Regression Analysis for Time Series，RATS)；实现神经网络和绘制附图用 S-Plus. 运行这些软件所需的某些命令在各章后的附录中给出. 特别地，用来估计多元波动率模型的复杂的 RATS 程序在第 9 章附录 A 中给出. 我采用了一些自己和其他人编的 Fortran 程序来对简单的期权定价、估计极值模型、计算风险值和贝叶斯(Bayesian)方法进行分析. 一些数据集和程序可以在国际互联网上获得，网址为：http://www.gsb.uchicago.edu/fac/ruey.tsay/teaching/fts.

第 1 章描述金融时间序列数据的一些基本特征. 其他章分为三个部分：第一部分由第 2 章至第 7 章组成，讨论一元金融时间序列的分析和应用；第二部分包括第 8 章和第 9 章，是关于多个资产收益率序列的；最后一部分是第 10 章，介绍用 MCMC 方法进行金融中的贝叶斯推断.

要完全读懂本书需要具备基本的统计学概念和知识. 在每章中，当一个必要的统计学概念第一次出现时，我都要给一个简短的回顾. 即使这样，还是竭力推荐统计学或商业统计学的必备知识，包括概率分布、线性回归分析. 读懂全书中所讨论的应用还需要了解金融知识. 然

而，具备较好计量经济学和统计学背景的读者，也会在本书中发现多方面有趣的主题和带挑战性的问题.

作为 MBA 的课程，第 2 章和第 3 章可作为核心内容，再加上一些非线性方法(如第 4 章的神经网络及第 5～7 章和第 10 章中讨论的应用). 对贝叶斯推断感兴趣的读者可以从第 10 章前 5 节开始阅读.

金融时间序列分析的研究发展迅速，新成果不断出现. 虽然我已经力图覆盖较广的范围，但仍有许多主题没有涉及或只是一带而过.

我真诚地感谢我的老师和亲密的朋友刁锦寰先生，是他在这些年中给了我指导、鼓励和钻研统计应用的坚定信念. 感谢 Steve Quigley、Heather Haselkorn、Leslie Galen、Danielle LaCourciere 和 Amy Hendrickson，没有他们的帮助这本书是不可能出版的. 感谢 Richard Smith，他送给我极值理论的估计程序. 感谢 Bonnie K. Ray，他对本书的好几章都有非常有益的建议. 感谢 Steve Kou 送给我他的关于跳跃扩散模型论文的预印本. 感谢 Robert E. McCulloch 许多年来在 MCMC 方法上的合作. 感谢选修我的金融时间序列分析课程的许多学生的反馈和投入. 感谢 Jeffrey Russell 和 Michael Zhang 关于高频金融数据的深入讨论. 也感谢芝加哥大学商学院和美国国家科学基金会的支持. 最后，对我的妻子 Teresa 的一贯支持、鼓励和理解，对 Julie、Richard 和 Vicki 给我带来的快乐和灵感，对我的父母亲给我的关爱，表示衷心的谢意.

R. S. T.

芝加哥，伊利诺伊州

目 录

第 1 章　金融时间序列及其特征

金融时间序列分析研究的是资产价值随时间演变的理论与实践．它是一个带有高度经验性的学科，但也像其他科学领域一样，理论是形成分析推断的基础．然而，金融时间序列分析有一个区别于其他时间序列分析的主要特点：金融理论及其经验的时间序列都包含不确定因素．例如，资产波动率有各种不同的定义，对一个股票收益率序列，波动率是不能直接观察到的．正因为带有不确定性，统计理论和方法在金融时间序列分析中起重要作用．

本书的目的是提供金融时间序列的一些知识，介绍一些对分析金融时间序列有用的统计工具，使读者获得各种经济计量方法在金融中应用的经验．第 1 章引入资产收益率的基本概念，并简短介绍全书的内容．第 2 章回顾一些线性时间序列分析中的基本概念，如平稳性、自相关函数，引入一些简单的模型来处理序列的先后相关性，讨论具有时间序列误差的回归模型、季节性、单位根非平稳性和长记忆过程等．第 3 章集中讨论条件异方差性(即资产收益率的条件方差)的建模，讨论各种新近发展起来的描述资产收益率的波动率随时间演变的各种经济计量模型．在第 4 章中我们讨论金融时间序列中的非线性性，引入能辨别非线性序列与线性序列的检验统计量，并讨论几个非线性模型．这一章还介绍非参数估计方法、神经网络，展示非线性模型在金融中的各种应用．第 5 章考虑高频金融数据的分析和这种分析在讨论市场微观结构中的应用，阐明不同步(或不同时)的交易和讨价还价的弹跳会产生股票收益的先后相关性，还研究不同交易之间时间持续的动态规律和一些分析交易数据的计量经济模型．在第 6 章中，我们引入连续时间扩散模型和动态规律的伊藤(Ito)引理，导出 Black-Scholes 期权定价公式，用简单的、带跳跃的扩散模型来刻画期权市场常见的一些特征．第 7 章讨论极值理论、厚尾分布及其在金融风险管理中的应用．特别地，要讨论计算一个金融头寸风险值(VaR)的各种方法．第 8 章集中讨论多元时间序列分析和简单的多元模型，研究时间序列之间的交叉延迟关系，讨论简化一个多元序列动态结构和降低维数的方法，引进协整(Co-integration)和门限协整(Threshold Co-integration)，并用来研究金融市场中的套利(Arbitrage)机会．第 9 章介绍多元波动率模型，其中包括带时变相关系数的模型，讨论怎样对一个条件协方差阵重新进行参数化，使之满足正定性的限制，并降低波动率建模的复杂性．最后，在第 10 章中，我们介绍统计文献中一些新近发展起来的蒙特卡罗马尔可夫链(MCMC)方法，并把这些方法应用于各种金融的研究问题，如随机波动率和马尔可夫转换模型的估计．

本书着重强调应用和经验数据分析，每章都有实际例子，很多时候是用金融时间序列的经验特征来作为产生经济计量模型的动机．必要时，还提供用来分析数据的计算程序和命令．有的程序在附录中给出．各章的练习题中用到很多实际数据集．

1.1 资产收益率

多数金融研究是针对资产收益率而不是资产价格. Campbell, Lo 和 MacKinlay(1997)给出了两个使用收益率的主要理由：第一，对普通的投资者来说，资产收益率是投资机会的完全的、尺度自由的概括；第二，收益率序列比价格序列更容易处理，因为前者有更好的统计性质. 然而，资产收益率有多种定义.

设 P_t 是资产在 t 时刻的价格. 下面给出全书中要用的一些收益率的定义. 暂时假定资产不支付分红.

单周期简单收益率

若从第 $t-1$ 天到第 t 天(一个周期)持有某种资产，则简单毛收益率为

$$1+R_t=\frac{P_t}{P_{t-1}} \quad 或 \quad P_t=P_{t-1}(1+R_t), \tag{1.1}$$

对应的单周期简单净收益率或称简单收益率为

2

$$R_t=\frac{P_t}{P_{t-1}}-1=\frac{P_t-P_{t-1}}{P_{t-1}}. \tag{1.2}$$

多周期简单收益率

若从第 $t-k$ 天到第 t 天这 k 个周期内持有某种资产，则 k 周期简单毛收益率为

$$\begin{aligned}1+R_t[k]&=\frac{P_t}{P_{t-k}}=\frac{P_t}{P_{t-1}}\times\frac{P_{t-1}}{P_{t-2}}\times\cdots\times\frac{P_{t-k+1}}{P_{t-k}}\\&=(1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1})\\&=\prod_{j=0}^{k-1}(1+R_{t-j}).\end{aligned}$$

这样，k 周期简单毛收益率就是所包括的 k 个单周期简单毛收益率之乘积，称为复合收益率. k 周期简单净收益率是 $R_t[k]=(P_t-P_{t-k})/P_{t-k}$.

在实践中，实际的时间区间对讨论和比较收益率很重要(例如是月收益率还是年收益率). 若时间区间没有给出，那么隐含假定时间区间为一年. 如果持有资产的期限为 k 年，则年度化的(平均的)收益率定义为

$$年度化的\{R_t[k]\}=\left[\prod_{j=0}^{k-1}(1+R_{t-j})\right]^{1/k}-1.$$

这是由所包括的 k 个单周期简单毛收益率的几何平均得到的，它可以用下式计算：

$$年度化的\{R_t[k]\}=\exp\left[\frac{1}{k}\sum_{j=0}^{k-1}\ln(1+R_{t-j})\right]-1,$$

其中 $\exp(x)$表示指数函数，$\ln(x)$是正数 x 的自然对数. 因为计算算术平均值比计算几何平均值容易，并且单周期收益率一般很小，我们可以用一阶泰勒(Taylor)展开去近似年度化的收益率，得到

$$年度化的\{R_t[k]\}\approx\frac{1}{k}\sum_{j=0}^{k-1}R_{t-j}. \tag{1.3}$$

然而，在有些场合，式(1.3)近似的程度并不好.

连续复合

在引进连续复合的收益率之前，我们讨论一下复合的效果. 假定银行存款的年利率为10%，最初存款为1美元. 假如该银行每年支付一次利息，那么一年之后存款的额度变为1+0.1=1.1美元. 假如该银行半年付息一次，六个月的利息率是10%/2=5%，第一年之后额度是$1(1+0.1/2)^2=1.1025$. 一般地，如果银行一年付息m次，那么每次支付的利息率为$10\%/m$，一年后存款的额度变成$1(1+0.1/m)^m$美元. 表1-1给出一些常用的时间间隔下年利率为10%时存款1美元的结果. 可见，净值趋于$1.1052\approx\exp(0.1)$，这个值就是连续复合的结果. 我们可以清楚地看到复合的效果.

3

表1-1 复合效果的说明：期限为1年，年利率为10%

类型	支付次数	每周期的利率	净值(美元)
一年	1	0.1	1.100 00
半年	2	0.05	1.102 50
季度	4	0.025	1.103 81
月	12	0.008 3	1.104 71
周	52	$\frac{0.1}{52}$	1.105 06
天	365	$\frac{0.1}{365}$	1.105 16
连续地	∞		1.105 17

一般地，连续复合的净资产值A是

$$A = C\exp(r\times n), \tag{1.4}$$

其中r是年利率，C是初始资本，n是年数(可为小数——译者注). 由式(1.4)，我们有

$$C = A\exp(-r\times n). \tag{1.5}$$

这叫做n年后价值为A的资产的现值，这里我们假定连续复合的年利率为r.

连续复合收益率

资产的简单毛收益率的自然对数称为连续复合收益率或对数收益率(log-return)：

$$r_t = \ln(1+R_t) = \ln\frac{P_t}{P_{t-1}} = p_t - p_{t-1}, \tag{1.6}$$

其中$p_t=\ln P_t$. 连续复合收益率r_t与简单净收益率R_t相比有一些优点. 首先，对多周期收益率，我们有

4

$$\begin{aligned} r_t[k] &= \ln(1+R_t[k]) = \ln[(1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1})] \\ &= \ln(1+R_t)+\ln(1+R_{t-1})+\cdots+\ln(1+R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1}. \end{aligned}$$

这样，连续复合多周期收益率就是所连续包括的单周期收益率之和. 其次，对数收益率有更容易处理的统计性质.

资产组合收益率

由N个资产组成的一个资产组合的简单净收益率是它所包含的各个资产的简单净收益率

的加权平均，其中每个资产的权重是资产组合的总价值中该资产的价值所占的百分比．设 p 是一个资产组合，其在资产 i 上的权重为 w_i，那么 p 在 t 时刻的简单收益率 $R_{p,t}=\sum_{i=1}^{N}w_iR_{it}$ ，其中 R_{it} 是资产 i 的简单收益率．

然而，资产组合的连续复合收益率没有上述方便的性质．如果简单收益率 R_{it} 的绝对值都很小的话，我们有 $r_{p,t}\approx\sum_{i=1}^{N}w_ir_{it}$ ，其中 $r_{p,t}$ 是该组合在 t 时刻的连续复合收益率．这种近似经常被用来研究资产组合的收益率．

分红

如果一个资产周期性地支付分红，我们必须修改资产收益率的定义．设 D_t 是一个资产在第 $t-1$ 天和第 t 天之间的分红，P_t 是该资产在第 t 个周期末的价格．这样，分红是不包含在 P_t 中的．因此，t 时刻简单净收益率和连续复合收益率变成

$$R_t=\frac{P_t+D_t}{P_{t-1}}-1,\quad r_t=\ln(P_t+D_t)-\ln P_{t-1}.$$

超出收益率

一个资产的超出收益率是该资产的收益率与某个参考资产的收益率之差．这个参考资产通常是无风险的，如美国短期国债的收益率．简单超出收益率和对数超出收益率分别定义为

$$Z_t=R_t-R_{0t},\quad z_t=r_t-r_{0t},\tag{1.7}$$

其中 R_{0t} 和 r_{0t} 分别是该参考资产的简单收益率和对数收益率．在金融文献中，超出收益率被认为是一个套利投资的赢利，在这个投资组合中对某资产持多头头寸而对参考资产持空头头寸．

注释： 多头金融头寸意指持有某资产．空头头寸意指卖出不属自己的资产．这需通过从已购买该资产的投资者那里借入资产来完成．在之后的某天，卖空者有义务买进和借入时完全相同数量的股份偿还给借出者．因为偿还时要求的是相等数量股份，而不是相等数量的美元，卖空者会由于该资产价格的下跌而获利．如果在空头持续期间对该资产有现金分红，则支付给做空买卖的买者．卖空者也必须从自己的资源里配备相应的现金分红来补偿借出者．换句话说，卖空者有义务支出所借资产的现金分红给借出者（见 Cox 和 Rubinstein (1985)）．

5

关系总结

简单收益率 R_t 与连续复合（或对数）收益率 r_t 的关系是

$$r_t=\ln(1+R_t),\quad R_t=e^{r_t}-1.$$

收益率的时间累积产生

$$1+R_t[k]=(1+R_t)(1+R_{t-1})\cdots(1+R_{t-k+1}),$$

$$r_t[k]=r_t+r_{t-1}+\cdots+r_{t-k+1}.$$

如果连续复合年利率为 r，则资产的现值与资产的将来值的关系为

$$A=C\exp(r\times n),\quad C=A\exp(-r\times n).$$

1.2　收益率的分布性质

要研究资产收益率，最好是从它们的分布性质开始．目的是弄清不同资产、不同时间收益率的表现．考虑 N 个资产，持有这 N 个资产 T 个时间周期，如 $t=1, \cdots, T$．对每个资产 i，r_{it} 表示它在 t 时刻的对数收益率．所要研究的对数收益率为 $\{r_{it}; i=1, \cdots, N; t=1, \cdots, T\}$．也可以考虑简单收益率 $\{R_{it}; i=1, \cdots, N; t=1, \cdots, T\}$ 和对数超出收益率 $\{z_{it}; i=1, \cdots, N; t=1, \cdots, T\}$．

1.2.1　统计分布及其矩的回顾

我们简短地回顾一下统计分布的一些基本性质和随机变量的矩．R^k 表示 k 维欧几里得空间，$\boldsymbol{x}\in R^k$ 表示 $\boldsymbol{x}$ 是 R^k 中的点，考虑两个随机向量 $\boldsymbol{X}=(X_1, \cdots, X_k)'$ 和 $\boldsymbol{Y}=(Y_1, \cdots, Y_q)'$．$P(\boldsymbol{X}\in A, \boldsymbol{Y}\in B)$ 表示 $\boldsymbol{X}$ 在子空间 $A\subset R^k$ 中、$\boldsymbol{Y}$ 在子空间 $B\subset R^q$ 中的概率．本书的大部分场合，都假定两个随机向量是连续的．　[6]

联合分布

函数

$$F_{X,Y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})=P(\boldsymbol{X}\leqslant \boldsymbol{x},\boldsymbol{Y}\leqslant \boldsymbol{y}),\quad \boldsymbol{x}\in R^p,\boldsymbol{y}\in R^q$$

是带参数 $\boldsymbol{\theta}$ 的 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的联合分布，其中不等号"$\leqslant$"是分量对分量的运算．$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的规律由 $F_{X,Y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$ 刻画．如果 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的联合概率密度函数 $f_{x,y}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$ 存在，则

$$F_{X,Y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{x,y}(\boldsymbol{w},\boldsymbol{z};\boldsymbol{\theta})\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{w}.$$

这时，$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是连续随机向量．

边际分布

$\boldsymbol{X}$ 的边际分布是

$$F_X(\boldsymbol{x};\boldsymbol{\theta})=F_{X,Y}(\boldsymbol{x},\infty,\cdots,\infty;\boldsymbol{\theta}).$$

这样，$\boldsymbol{X}$ 的边际分布可通过对 $\boldsymbol{Y}$ 求和(求积分)得到．对 $\boldsymbol{Y}$ 的边际分布也可类似得到．

如果 $k=1$，X 是一个一元随机变量，其分布函数为

$$F_X(x)=P(X\leqslant x;\boldsymbol{\theta}),$$

称为 X 的积累分布函数(cumulative distribution function，CDF)．一个随机变量的 CDF 是不减的[即，对 $x_1\leqslant x_2$ 有 $F_X(x_1)\leqslant F_X(x_2)$，且 $F_X(-\infty)=0$，$F_X(\infty)=1$]．对给定的概率 p，使 $p\leqslant F_X(x_p)$ 的最小实数 x_p 叫做随机变量 X 的 p 分位点，即

$$x_p=\inf_x\{x\,|\,p\leqslant F_X(x)\}.$$

条件分布

给定 $\boldsymbol{Y}\leqslant \boldsymbol{y}$ 的条件下 $\boldsymbol{X}$ 的条件分布为

$$F_{X|Y\leqslant y}(\boldsymbol{x};\boldsymbol{\theta})=\frac{P(\boldsymbol{X}\leqslant \boldsymbol{x},\boldsymbol{Y}\leqslant \boldsymbol{y})}{P(\boldsymbol{Y}\leqslant \boldsymbol{y})}.$$

若所对应的概率密度函数存在，则给定 $\boldsymbol{Y}=\boldsymbol{y}$ 的条件下，$\boldsymbol{X}$ 的条件密度为　[7]

$$f_{x|y}(\boldsymbol{x};\boldsymbol{\theta})=\frac{f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})}{f_y(\boldsymbol{y};\boldsymbol{\theta})},\tag{1.8}$$

其中边际密度函数 $f_y(\boldsymbol{y};\ \boldsymbol{\theta})$ 由下式得到

$$f_y(\boldsymbol{y};\boldsymbol{\theta})=\int_{-\infty}^{\infty}f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})\mathrm{d}\boldsymbol{x}.$$

从式(1.8)看，联合分布、边际分布和条件分布之间的关系为

$$f_{x,y}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})=f_{x|y}(\boldsymbol{x};\boldsymbol{\theta})\times f_y(\boldsymbol{y};\boldsymbol{\theta}).\tag{1.9}$$

这个相等关系在时间序列分析中经常用到(如在进行最大似然估计时). 最后，$\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 是相互独立的随机向量，当且仅当 $f_{x|y}(\boldsymbol{x};\ \boldsymbol{\theta})=f_x(\boldsymbol{x};\ \boldsymbol{\theta})$，这时 $f_{x,y}(\boldsymbol{x},\ \boldsymbol{y};\ \boldsymbol{\theta})=f_x(\boldsymbol{x};\ \boldsymbol{\theta})f_y(\boldsymbol{y};\ \boldsymbol{\theta})$.

随机变量的矩

一个连续型随机变量 X 的 l 阶矩定义为

$$m'_l=E(X^l)=\int_{-\infty}^{\infty}x^lf(x)\mathrm{d}x,$$

其中"E"表示期望，$f(x)$ 是 X 的概率密度函数. 一阶矩称为 X 的均值(mean)或期望，它表示的是分布的中心位置，记为 μ_x. X 的 l 阶中心矩定义为

$$m_l=E[(X-\mu_x)^l]=\int_{-\infty}^{\infty}(x-\mu_x)^lf(x)\mathrm{d}x,$$

只要式中的积分是存在的. 二阶中心矩可度量 X 取值的变化程度，称为 X 的*方差*(variance)，记为 σ_x^2. 方差的正平方根 σ_x 称为 X 的*标准差*. 一个正态分布是由随机变量的头两阶矩唯一决定的. 对其他分布，可能要了解其更高阶矩.

三阶中心矩度量 X 关于其均值的对称性，而四阶中心矩度量 X 的尾部，在统计学中，标准化的三阶矩叫*偏度*(skewness)，标准化的四阶矩叫做*峰度*(kurtosis)，它们分别用来描述随机变量的对称程度和尾部厚度. 具体地，X 的偏度和峰度定义为

[8]

$$S(x)=E\left[\frac{(X-\mu_x)^3}{\sigma_x^3}\right],\quad K(x)=E\left[\frac{(X-\mu_x)^4}{\sigma_x^4}\right].$$

量 $K(x)-3$ 叫做*超出峰度*(excess kurtosis)，因为对一个正态分布，$K(x)=3$. 这样，正态随机变量的超出峰度为 0. 具有正的超出峰度的分布称具有厚尾性，意指该分布在其支撑(support)的尾部有比正态分布更多的"质量". 在实际中，这意味着来自于这样一个分布的随机样本会有更多的极端值.

在应用中，偏度和峰度可以由它们对应的样本偏度和样本峰度来估计. 设 $\{x_1,\ \cdots,\ x_T\}$ 是 X 的 T 个观察值的随机样本，样本均值为

$$\hat{\mu}_x=\frac{1}{T}\sum_{t=1}^{T}x_t,\tag{1.10}$$

样本方差为

$$\hat{\sigma}_x^2=\frac{1}{T-1}\sum_{t=1}^{T}(x_t-\hat{\mu}_x)^2,\tag{1.11}$$

样本偏度为

$$\hat{S}(x)=\frac{1}{(T-1)\hat{\sigma}_x^3}\sum_{t=1}^{T}(x_t-\hat{\mu}_x)^3, \tag{1.12}$$

样本峰度为

$$\hat{K}(x)-3=\frac{1}{(T-1)\hat{\sigma}_x^4}\sum_{t=1}^{T}(x_t-\hat{\mu}_x)^4. \tag{1.13}$$

在正态分布假定下，$\hat{S}(x)$和$\hat{K}(x)$均渐近正态分布，均值为零，而方差分别是 $6/T$ 和 $24/T$，(见 Snedecor 和 Cochran(1980)，第 78 页).

1.2.2 收益率的分布

对数收益率$\{r_{it};\ i=1,\ \cdots,\ N;\ t=1,\ \cdots,\ T\}$的最一般的模型是它们的联合分布函数：

$$F_r(r_{11},\cdots,r_{N1};r_{12},\cdots,r_{N2};\cdots;r_{1T},\cdots,r_{NT};\boldsymbol{Y};\boldsymbol{\theta}), \tag{1.14}$$

其中$\boldsymbol{Y}$是由一些变量组成的状态向量，这些变量描述了决定资产收益率的环境，$\boldsymbol{\theta}$是唯一决定分布函数$F_r(.)$的参数向量. 概率分布$F_r(.)$刻画了收益率r_{it}和$\boldsymbol{Y}$的随机行为. 在许多金融研究中，状态向量$\boldsymbol{Y}$当作给定的，而主要关心的是给定$\boldsymbol{Y}$的条件下$\{r_{it}\}$的条件分布. 因此，资产收益率的实证分析是去估计未知参数$\boldsymbol{\theta}$，并在给定一些过去的对数收益率的条件下，对$\{r_{it}\}$的行为做出统计推断.

9

式(1.14)的模型太广了，以至于没有实际应用价值. 然而，它提供了一个一般的框架，在这个框架中，可建立资产收益率r_{it}的适当的计量经济模型.

有些金融理论，比如 Sharpe 在 1964 年提出的资本资产定价模型(Capital Asset Pricing Model，CAPM)，考虑的是在单个时间点t上N个收益率的联合分布(即$\{r_{1t},\ \cdots,\ r_{Nt}\}$的分布). 另外一些理论则强调各个资产收益率的动态结构(即对一个给定的资产i，$\{r_{i1},\ \cdots,\ r_{iT}\}$的分布). 本书中，我们对这两个方面都很关心. 在第 2 章至第 7 章的一元分析中，主要关心的是对资产i，$\{r_{it}\}_{t=1}^{T}$的联合分布. 为此，把联合分布分解成如下形式

$$\begin{aligned}F(r_{i1},\cdots,r_{iT};\boldsymbol{\theta})&=F(r_{i1})F(r_{i2}\mid r_{1t})\cdots F(r_{iT}\mid r_{i,T-1},\cdots,r_{i1})\\&=F(r_{i1})\prod_{t=2}^{T}F(r_{it}\mid r_{i,t-1},\cdots,r_{i1}).\end{aligned} \tag{1.15}$$

这个分解式突出了对数收益率r_{it}在时间上的先后相依性. 因此，主要问题就是条件分布$F(r_{it}\mid r_{i,t-1},\ \cdot)$的具体化；特别是条件分布是怎样随时间演变的. 在金融中，不同的分布的具体化会导出不同的理论. 例如，随机游动假定的一种形式就是条件分布$F(r_{it}\mid r_{i,t-1},\ \cdots,\ r_{i1})$等于边际分布$F(r_{it})$. 这时，收益率在时间上是先后相互独立的，从而是不可预报的.

通常地，我们把资产收益率当作连续型随机变量对待，尤其是对低频计算出的指数收益率或股票收益率，用它们的概率密度函数. 这时，利用等式(1.9)，我们把式(1.15)的分解写成

$$f(r_{i1},\cdots,r_{iT};\boldsymbol{\theta})=f(r_{i1};\boldsymbol{\theta})\prod_{t=2}^{T}f(r_{it}\mid r_{i,t-1},\cdots,r_{i1},\boldsymbol{\theta}). \tag{1.16}$$

对高频资产收益率，离散性就变成一个问题. 例如，在纽约股票交易所(New York Stock

Exchange，NYSE)股票的价格是以一个微小量(tick size)的倍数变化的．这个微小量在1997年7月之前取为1/8美元，而在1997年7月至2001年1月是1/16美元．因此，NYSE记录的个股的收益率不是连续型的．我们将在第5章讨论高频的股价变化和在价格发生变化之间的时间持续期．

[10]

注释：2000年8月28日，纽约股票交易所开始了一项试验性的程序，对七支股票以十进制小数计价，而美国股票交易所(American Stock Exchange，AMEX)开始对六支股票和两种期权种类以十进制小数计价．在2000年9月25日和12月4日，NYSE分别有57个股票和94个股票加入该程序．在2001年1月29日，所有在NYSE和AMEX交易的股票都开始以十进制进行交易．

式(1.16)启示我们在资产收益率的研究中条件分布比边际分布更常用．然而，边际分布仍然是需要关心的．特别是，利用过去的收益率数据，估计边际分布比估计条件分布容易．另外，有时通过实证看出资产收益率只有很弱的先后相关性，从而它们的边际分布与条件分布是很相近的．

在对资产收益率的边际分布研究的文献中，已经用到好几种统计分布，包括正态分布、对数正态分布、稳定分布和正态分布的尺度混合(scale-mixture)．下面我们简短讨论一下这几种分布．

正态分布

金融研究中传统的假设是：简单收益率$\{R_{it}|t=1,\cdots,T\}$是独立的，且都服从一个固定均值和方差的正态分布．这个假设使得资产收益率的统计性质变得可以处理，但它遇到几个麻烦：第一，简单资产收益率的下界为-1，而正态分布的支撑是没有下界的，实直线上的任何值都可能取到；第二，如果R_{it}是正态分布的，那么多周期的简单收益率$R_{it}[k]$就不是正态分布的，因为它是单周期收益率的乘积；第三，经验结果不支持正态性假设，很多资产收益率数据表明它具有正的超出峰度．

对数正态分布

另一个常用的假定是：资产的对数收益率r_t是相互独立的且都服从均值为μ，方差为σ^2的正态分布．那么，简单收益率是独立同分布的对数正态的随机变量，均值和方差分别为

$$E(R_t)=\exp\left(\mu+\frac{\sigma^2}{2}\right)-1,\quad \mathrm{Var}(R_t)=\exp(2\mu+\sigma^2)[\exp(\sigma^2)-1].\tag{1.17}$$

这两个式子在研究资产收益率时是有用的(如用给对数收益率所建立的模型进行预报时)．反之，假设简单收益率服从对数正态分布，均值为m_1，方差为m_2，则对应的对数收益率r_t的均值和方差分别为

$$E(r_t)=\ln\left[\frac{m_1+1}{\sqrt{1+\frac{m_2}{(1+m_1)^2}}}\right],\quad \mathrm{Var}(r_t)=\ln\left[1+\frac{m_2}{(1+m_1)^2}\right].$$

因为有限个独立同分布的正态随机变量之和仍服从正态分布，在$\{r_t\}$的正态假定下$r_t[k]$

[11]

也是正态的．另外，对r_t没有下界，并且从$1+R_t=\exp\{r_t\}$可见R_t的下界也能满足．然而，对数正态假定不是与历史股票收益率的所有性质都一致的，特别是很多股票收益率表现出正的超出峰度．

稳定分布

稳定分布是正态分布的自然推广，它们在加法运算下是稳定的，这一点符合连续复合收益率 r_t 的要求. 另外，稳定分布能刻画历史股票收益率所显示出来的超出峰度. 然而，非正态稳定分布没有有限方差，这一点与大部分金融理论相矛盾. 还有，用非正态的稳定分布进行统计建模是困难的. 非正态稳定分布的一个例子是柯西(Cauchy)分布，它关于它的中位数是对称的，但方差是无限的.

正态分布的尺度混合

股票收益率的新近研究中，倾向于利用正态分布的尺度混合或有限混合. 在正态分布的尺度混合的假定下，对数收益率 r_t 是服从均值为 μ、方差为 σ^2 的正态分布(即 $r_t \sim N(\mu, \sigma^2)$)，但是，$\sigma^2$ 是一个随机变量，它服从一个正的分布(如 σ^{-2}服从一个伽马(Gamma)分布). 正态分布的有限混合的一个例子是

$$r_t \sim (1-X)N(\mu,\sigma_1^2)+XN(\mu,\sigma_2^2),$$

其中 X 是伯努利随机变量，$P(X=1)=\alpha$，$0\leqslant\alpha\leqslant 1$，$\sigma_1^2$ 较小而 σ_2^2 较大. 例如，对 $\alpha=0.05$，有限混合指的是 95%的收益率服从$N(\mu, \sigma_1^2)$，5%的收益率服从 $N(\mu, \sigma_2^2)$，σ_2^2 的较大值使混合把更多的“质量”放在分布的尾部. 来自于 $N(\mu, \sigma_2^2)$的收益率的百分比较低，说的是大多数收益率服从一个简单的正态分布. 正态分布混合的优点包括它们保持了正态分布的易操作性、有有限高阶矩，能抓住超出峰度. 但难以估计混合参数(如有限混合中的 α).

图 1-1 显示的是正态分布有限混合、柯西分布和标准正态分布的概率密度函数. 正态有限混合是$(1-X)N(0, 1)+XN(0, 16)$，$\alpha=0.05$，混合密度是标准的. 柯西密度函数是

$$f(x)=\frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty.$$

可见，柯西分布有比正态有限混合更厚的尾部，而正态有限混合有比标准正态更厚的尾部. [12]

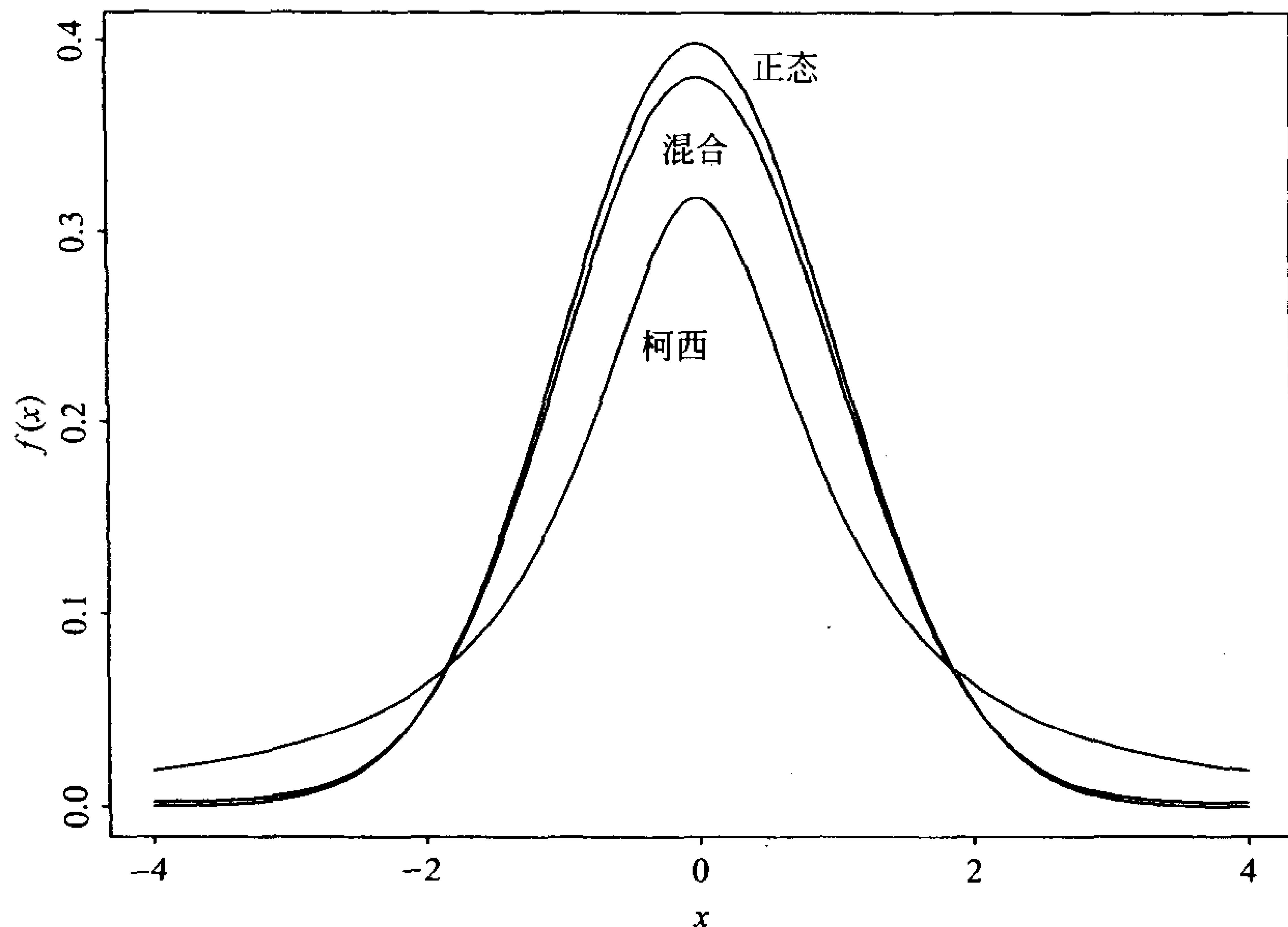


图 1-1 有限混合分布、稳定分布和标准正态分布密度函数的比较

1.2.3 多元收益率

设 $\boldsymbol{r}_t=(r_{1t}, \cdots, r_{Nt})'$是 N 个资产在 t 时刻的对数收益率. 第 8 章和第 9 章的多元分析将涉及$\{\boldsymbol{r}_t\}_{t=1}^T$的联合分布. 这个联合分布可以与式(1.15)同样的方式分解，因此分析集中在条件分布函数 $F(\boldsymbol{r}_t|\boldsymbol{r}_{t-1},\cdots,\boldsymbol{r}_1, \boldsymbol{\theta})$ 的具体化上. 特别是，r_t 的条件期望和条件协方差阵怎样随时间演变构成第 8 章和第 9 章的主题.

随机向量 $\boldsymbol{X}=(X_1, \cdots, X_p)$的均值向量和协方差矩阵定义为

$$E(\boldsymbol{X})=\boldsymbol{\mu}_x=[E(X_1), \cdots, E(X_p)]',$$
$$\mathrm{Cov}(\boldsymbol{X})=\Sigma_x=E[(\boldsymbol{X}-\boldsymbol{\mu}_x)(\boldsymbol{X}-\boldsymbol{\mu}_x)'],$$

只要所涉及到的期望是存在的. 当有来自于 $\boldsymbol{X}$ 的数据$\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T\}$时，样本的均值和样本协方差阵定义为

[13]

$$\hat{\boldsymbol{\mu}}_x=\frac{1}{T}\sum_{t=1}^T\boldsymbol{x}_t,\ \hat{\Sigma}_x=\frac{1}{T}\sum_{t=1}^T(\boldsymbol{x}_t-\hat{\boldsymbol{\mu}}_x)(\boldsymbol{x}_t-\hat{\boldsymbol{\mu}}_x)'.$$

只要 $\boldsymbol{X}$ 的协方差阵存在，这些样本的统计量都是它们对应的理论值的相合估计. 在金融文献中多元正态分布常用来描述对数收益率 $\boldsymbol{r}_t$.

1.2.4 收益率的似然函数

式(1.15)的分解可用来得到一个资产的对数收益率$\{r_1, \cdots, r_T\}$的似然函数. 这里为了符号上的简便，对数收益率中的下角标 i 省略不写. 若条件分布$f(r_t|r_{t-1}, \cdots, r_1, \boldsymbol{\theta})$是均值为$\mu_t$、方差为 σ_t^2 的正态分布，则$\boldsymbol{\theta}$由参数μ_t 和σ_t^2 组成，数据的似然函数为

$$f(r_1,\cdots,r_T;\boldsymbol{\theta})=f(r_1;\boldsymbol{\theta})\prod_{t=2}^T\frac{1}{\sqrt{2\pi}\sigma_t}\exp\left[-\frac{(r_t-\mu_t)^2}{2\sigma_t^2}\right], \tag{1.18}$$

其中 $f(r_1;\boldsymbol{\theta})$是第一个观察$r_1$ 的边际密度函数，使似然函数达到最大值的$\boldsymbol{\theta}$的值就是参数$\boldsymbol{\theta}$ 的最大似然估计(MLE). 因为对数函数是单调的，MLE 可通过最大化对数似然函数得到:

$$\ln f(r_1,\cdots,r_T;\boldsymbol{\theta})=\ln f(r_1;\boldsymbol{\theta})-\frac{1}{2}\sum_{t=2}^T\left[\ln(2\pi)+\ln(\sigma_t^2)+\frac{(r_t-\mu_t)^2}{\sigma_t^2}\right].$$

对数似然函数在实际中更容易处理一些. 若条件分布 $f(r_t|r_{t-1},\cdots, r_1, \boldsymbol{\theta})$不是正态的，数据的对数似然函数可用相似的方式得到.

1.2.5 收益率的经验性质

这一节所用数据是从芝加哥大学证券价格研究中心(CRSP)得到的. 如果有分红支付，也包含在收益率之中了. 图 1-2 显示的是 IBM(International Business Machines)股票从 1926 年 1 月到 1997 年 12 月，每月简单收益率和对数收益率的时间图(time plot). 时间图就是数据对时间指标所成图像. 上方的图是简单收益率的图. 图 1-3 显示的是价值加权(value-weighted)市场指数的月收益率的同样的时间图. 如想像的一样，图像表明简单收益率和对数收益率的基本模式相似.

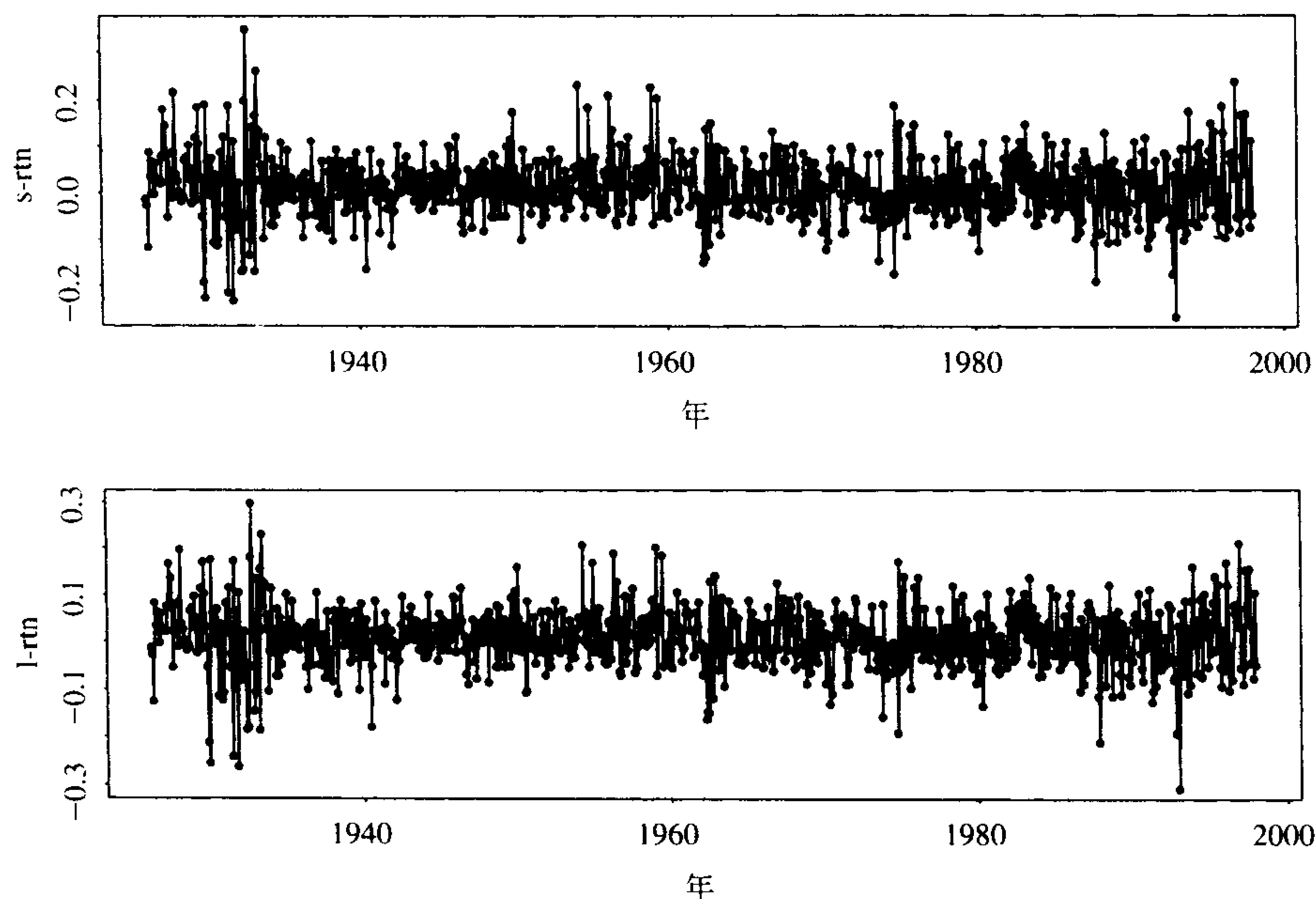


图 1-2 IBM 股票从 1926 年 1 月到 1997 年 12 月每月收益率的时间图．上方的图是简单净收益率，下方的图是对数收益率

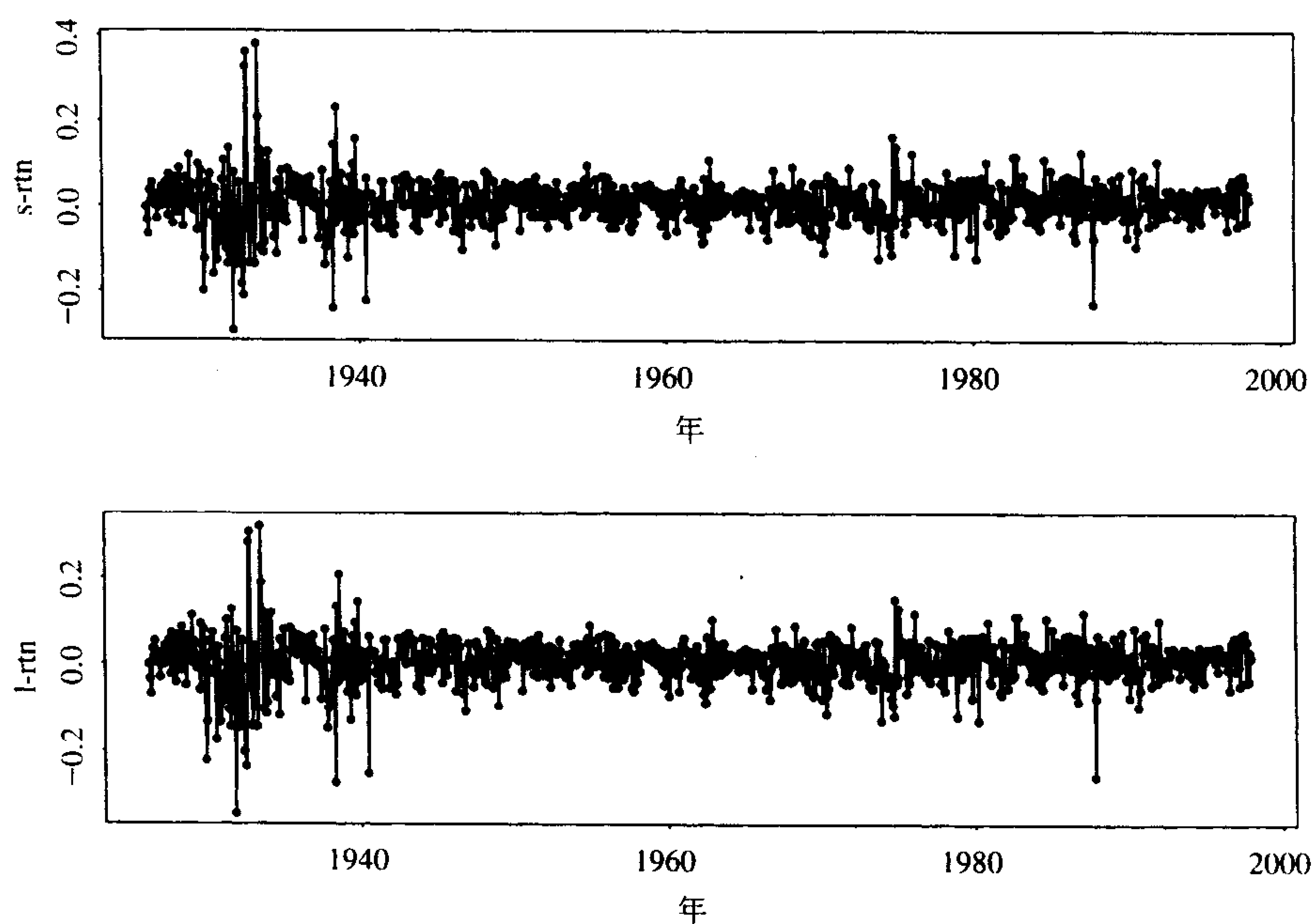


图 1-3 价值加权市场指数从 1926 年 1 月到 1997 年 12 月每月收益率的时间图．上方的图是简单净收益率，下方的图是对数收益率

表 1-2 列出了所选的美国股票指数和个股的简单收益率和对数收益率的一些描述性统计量的值. 所列收益率是日收益率和月收益率，它们的值是以百分比给出的. 所用数据的起始时间和样本容量大小也在表中给出. 从这个表上我们观察到如下几点：(a)股市指数和个股的日收益率具有很高的超出峰度，而对月收益率序列，股指的月收益率的超出峰度比个股的月收益率的超出峰度高出许多；(b)日收益率的均值接近于零，而月收益率的均值要稍大一些；(c)月收益率的标准差比日收益率的标准差大；(d)在日收益率中，股指的标准差比个股的标准差小，这一点与通常的感觉是相符的；(e)偏度不是一个严重问题，这一点对日收益率和月收益率都是一样；(f)描述性统计量表明简单收益率和对数收益率的差别是很小的.

14～15

表 1-2

证券	起始时间	样本容量	均值	标准差	偏度	超出峰度	最小值	最大值
(a) 日简单收益率(%)								
VW	62/7/3	8 938	0.049	0.798	−1.23	30.06	−17.18	8.67
EW	62/7/3	8 938	0.083	0.674	−1.09	18.09	−10.48	6.95
I. B. M.	62/7/3	8 938	0.050	1.479	0.01	11.34	−22.96	12.94
Intel	72/12/15	6 329	0.138	2.880	−0.17	6.76	−29.57	26.38
3M	62/7/3	8 938	0.051	1.395	−0.55	16.92	−25.98	11.54
Microsoft	86/3/14	2 985	0.201	2.422	−0.47	12.08	−30.13	17.97
Citi-Grp	86/10/30	2 825	0.125	2.124	−0.06	9.16	−21.74	20.75
(b)日对数收益率(%)								
VW	62/7/3	8 938	0.046	0.803	−1.66	40.06	−18.84	8.31
EW	62/7/3	8 938	0.080	0.676	−1.29	19.98	−11.08	6.72
I. B. M.	62/7/3	8 938	0.039	1.481	−0.33	15.21	−26.09	12.17
Intel	72/12/15	6 329	0.096	2.894	−0.59	8.81	−35.06	23.41
3M	62/7/3	8 938	0.041	1.403	−1.05	27.03	−30.08	10.92
Microsoft	86/3/14	2 985	0.171	2.443	−1.10	19.65	−35.83	16.53
Citi-Grp	86/10/30	2 825	0.102	2.128	−0.44	10.68	−24.51	18.86
(c)月简单收益率(%)								
VW	26/1	864	0.99	5.49	0.23	8.13	−29.00	38.28
EW	26/1	864	1.32	7.54	1.65	15.24	−31.23	65.51
I. B. M.	26/1	864	1.42	6.70	0.17	1.94	−26.19	35.12
Intel	72/12	300	2.86	12.95	0.59	3.29	−44.87	62.50
3M	46/2	623	1.36	6.46	0.16	0.89	−27.83	25.77
Microsoft	86/4	141	4.26	10.96	0.81	2.32	−24.91	51.55
Citi-Grp	86/11	134	2.55	9.17	−0.14	0.47	−26.46	26.08
(d)月对数收益率(%)								
VW	26/1	864	0.83	5.48	−0.53	7.31	−34.25	32.41
EW	26/1	864	1.04	7.24	0.34	8.91	−37.44	50.38
I. B. M.	26/1	864	1.19	6.63	−0.22	2.05	−30.37	30.10
Intel	72/12	300	2.03	12.63	−0.32	3.20	−59.54	48.55
3M	46/2	623	1.15	6.39	−0.14	1.32	−32.61	22.92
Microsoft	86/4	141	3.64	10.29	0.29	1.32	−28.64	41.58
Citi-Grp	86/11	134	2.11	9.11	−0.50	1.14	−30.73	23.18

注：股票指数和个股的每天和每月简单收益率和对数收益率的描述性统计量. 收益率的值是百分比，样本的终止时间为 1997 年 12 月 31 日，统计量是由式(1.10)～式(1.13)定义的，而 VW 和 EW 分别表示价值权重和相等权重两个指数.

图 1-4 显示的是 IBM 股票的月简单收益率和对数收益率的经验密度函数. 虚线是由表 1-2 中IBM 股票收益率的样本均值和样本标准差决定的正态概率密度函数. 图像显示出对 IBM 股票的月收益率的正态性假定是有疑问的. 经验密度函数与对应的正态分布密度相比，在均值附近有更高的峰，但尾部更厚. 换句话说，经验密度函数更高、更瘦，但有更宽的支撑.

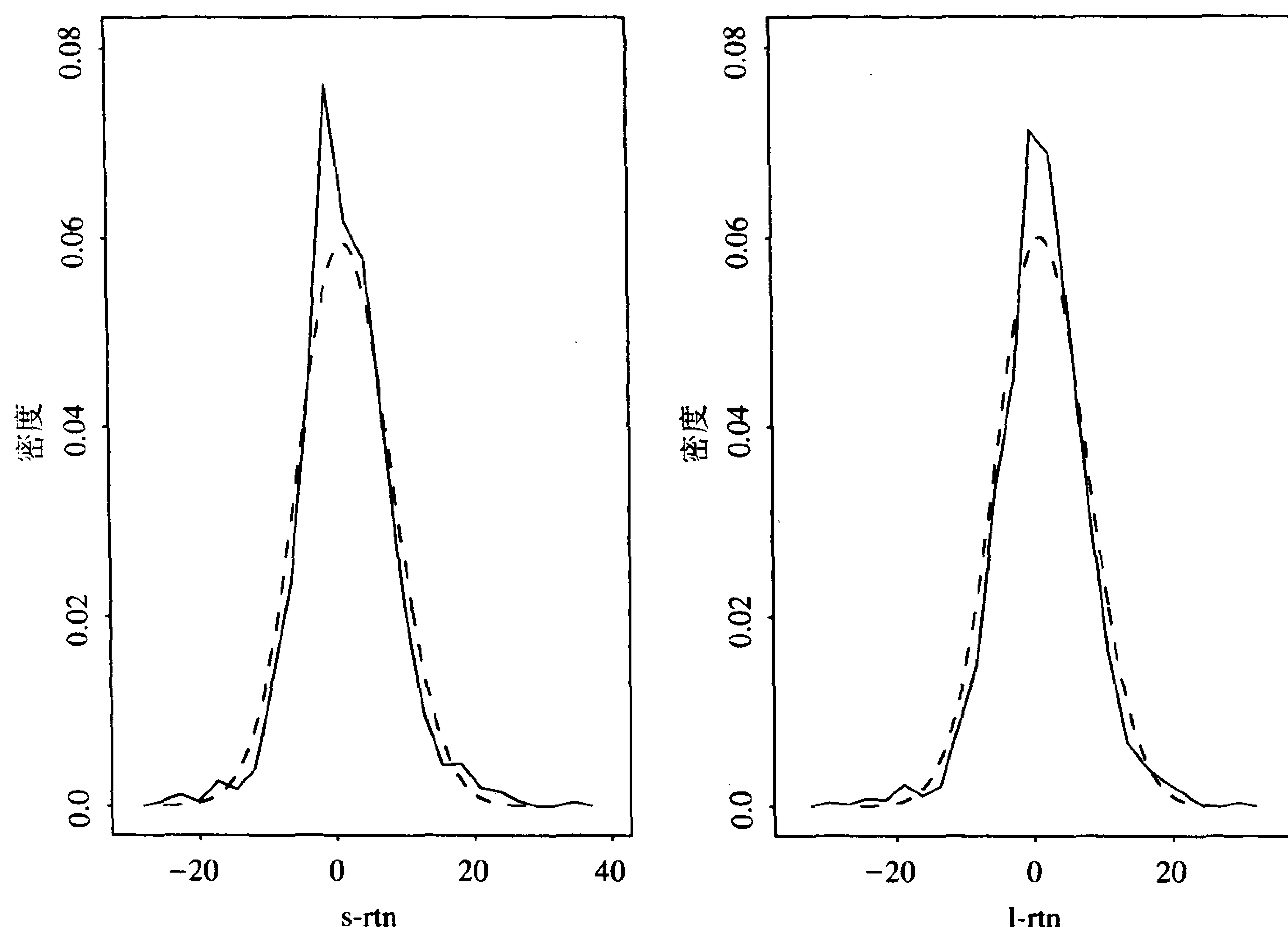


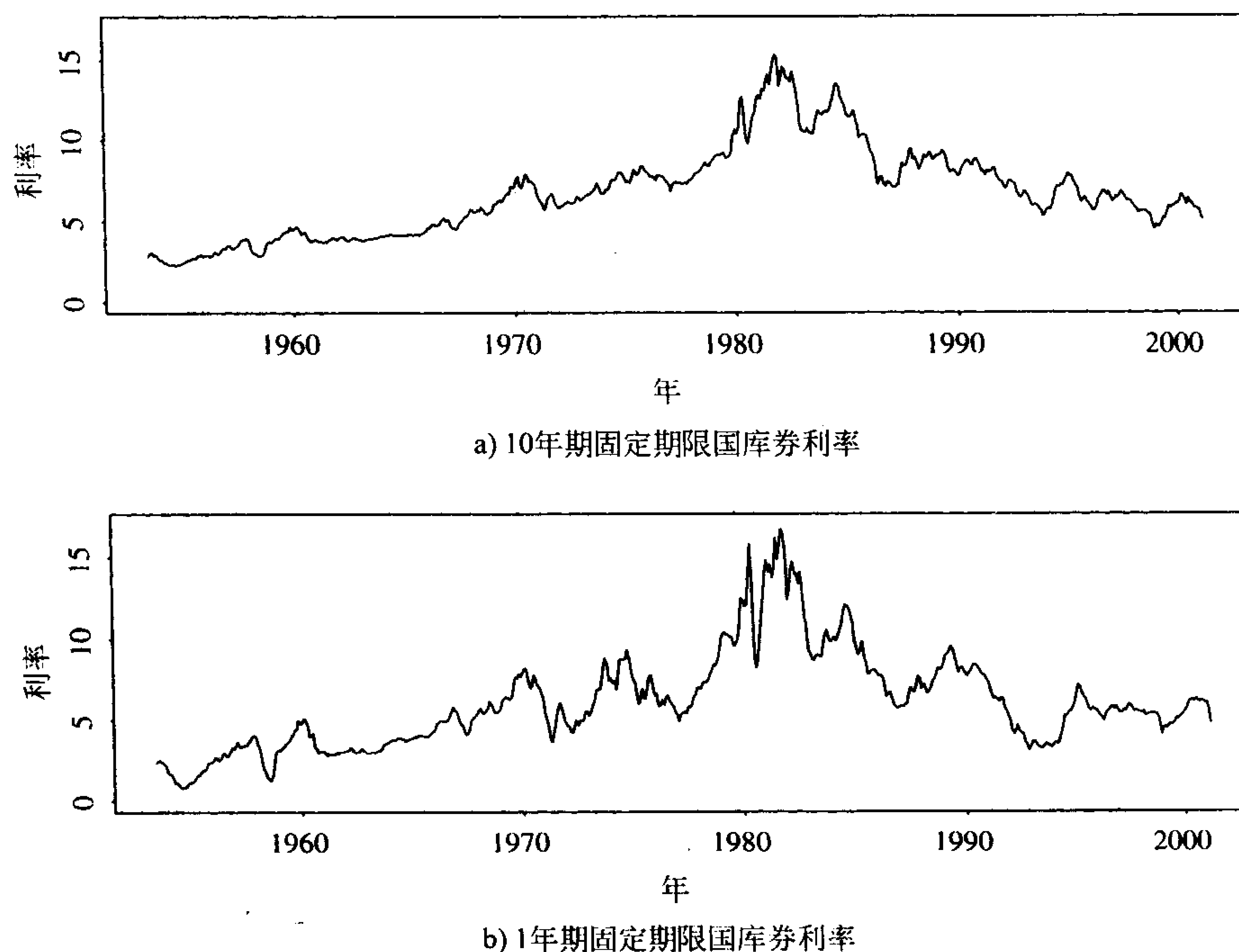
图 1-4 IBM 股票的月简单收益率和对数收益率的经验密度函数与正态密度函数的比较. 样本时间段是从 1926 年 1 月到 1997 年 12 月，左图是简单收益率，右图是对数收益率. 虚线是由表 1-2 给出的样本均值和样本标准差决定的正态分布密度函数

1.3 其他过程

除了收益率序列以外，我们还考虑波动率过程和资产极值收益率的行为. 波动率过程研究的是收益率的条件方差随时间演变的规律. 这是一个令人感兴趣的问题，因为如图 1-2 和图 1-3 所示的那样，收益率变化的大小是随时间变化的，且呈现出聚类现象. 在应用中，波动率在股票期权的定价中起重要作用. 收益率序列的极值指的是绝对值大的正或负的收益率. 表 1-2 说明收益率序列的最大值和最小值是不可忽视的. 负的极值收益率在风险管理中很重要，而正的极值收益率对持有空头是至关重要的. 我们将在第 7 章中研究极值收益率的性质和应用，例如极值发生的频率、极值的大小和经济变量对极值的影响等.

16
~
17

本书中考虑的其他金融时间序列包括利率、汇率、债券收益率、公司的每个股份在每季度的赢利. 图 1-5 所示的是美国的两种月利率的时间图，分别是从 1954 年 4 月至 2001 年 1 月的 10 年期和 1 年期具有固定期限的国库券利率. 正如我们所想像的，这两种利率的变化基本同步，但 1 年期利率的波动程度要更大一些. 表 1-3 提供了一些美国金融时间序列的描述性统计量的值，其中债券的月收益率是从 CRSP 得到的，时间是从 1942 年 1 月到 1999 年 12 月，利率数据是从圣・路易斯联邦储备银行(Federal Reserve Bank of St Louis)得到的，3 个月国库券的周利率是从 1954 年 1 月 8 日开始，6 个月国库券的周利率是从 1958 年 12 月 12 日开始，这两个序列的截止日期都是 2001 年 2 月 16 日. 对利率序列，期限越长样本均值越大，但样本标准差越小. 对债券收益率序列，样本标准差与期限长度正相关，而样本均值对不同期限保持稳定. 所考虑的大部分序列有正的超出峰度.



a) 10年期固定期限国库券利率

b) 1年期固定期限国库券利率

图 1-5 1954 年 4 月到 2001 年 1 月美国的月利率的时间图

关于表 1-2 所示的收益率的经验特征，第 2 章到第 4 章的内容涉及收益率序别头四阶矩，第 7 章讨论收益率最小值和最大值的行为，而第 8 章和第 9 章研究多个资产收益率的矩及它们之间的关系. 第 5 章讨论当时间区间很小时资产收益率的性质，第 6 章介绍数理金融的基本内容.

表 1-3

期限	均值	标准差	偏度	超出峰度	最小值	最大值
(a)债券的月收益率：1942 年 1 月至 1999 年 12 月，$T=696$						
30 年	0.43	2.53	0.66	2.77	−7.73	13.31
20 年	0.45	2.43	0.79	4.08	−8.41	15.24
10 年	0.45	1.97	0.71	2.72	−6.67	10.00
5 年	0.46	1.39	0.87	6.61	−5.80	10.61
1 年	0.44	0.53	2.50	16.72	−1.72	5.61
(b)国库券月利率：1953 年 4 月至 2001 年 1 月，$T=574$						
10 年	6.74	2.75	0.74	0.30	2.29	15.32
5 年	6.59	2.78	0.83	0.63	1.85	15.93
3 年	6.43	2.82	0.89	0.85	1.47	16.22
1 年	6.05	2.93	1.01	1.29	0.82	16.72
(c)国库券周利率：截止 2001 年 2 月 16 日						
6 个月	6.08	2.56	1.26	1.82	2.35	15.76
3 个月	5.51	2.76	1.14	1.88	0.58	16.76

注：某些美国金融时间序列的描述性统计量．数据的值是百分比．3 个月国库券的周利率起始时间是 1954 年 1 月 8 日，6 个月国库券起始时间为 1958 年 12 月 12 日．

练习题

1. 考虑如下公司股票从 1990 年 1 月到 1999 年 12 月每天收益率各 2 528 个观察值：Alcoa(aa)，American Express(axp)，Walt Disney(dis)，Chicago Tribune(trb) 和 Tyco International (tyc)．你可以直接从 CRSP 或从网上的文件获得这些数据．原始数据是来自于 CRSP 的持有期收益率，网上文件中的数据是已经转换为对数收益率并以百分比给出的．我们用股票账目记号来作为文件的名称(如“d-aa9099.dat”包含的是 Alcoa 公司股票从 1990 年到 1999 年的日对数收益率)．

   18～19

   - 计算日对数收益率的样本均值、方差、偏度、超出峰度、最小值、最大值．
   - 把对数收益率转换成简单收益率．计算日简单收益率的样本均值、方差、偏度、超出峰度、最小值、最大值．
   - 对数收益率的样本均值是显著地不同于零吗？在 5%的显著水平下得出你的结论，并讨论实际含义．

2. 考虑 Alcoa(aa)，General Motors(gm)，Walt Disney(dis) 和 Hershey Foods(hsy) 这四个公司股票的从 1962 年 1 月到 1999 年 12 月的月收益率各 456 个观察值，和 American Express (axp)、Mellon Financial Corporation(mel)两个公司股票的从 1973 年 1 月到 1999 年 12 月的月收益率各 324 个观察值．你可以直接从 CRSP 或从网上的文件中得出这些数据．用股票账目记号和所包含的年份作为数据文件的名称(如“m-mel7399.dat”包含的是 Melton Financial Corporation 公司股票从 1973 年 1 月到 1999 年 12 月的月对数收益率，以百分比给出)．

   - 计算月对数收益率的样本均值、方差、偏度、超出峰度，最小值和最大值．

- 把对数收益率转换成简单收益率. 计算月简单收益率的样本均值、方差、偏度、超出峰度、最小值、最大值.
- 对数收益率的样本均值是显著地不同于零吗? 在 5%的显著水平下得出你的结论，并讨论其实际含义.

3. 考虑 Alcoa 公司股票从 1962 年到 1999 年的月收益率:
   - 在给定数据的起止时间段上，平均的年对数收益率是多少?
   - 在给定数据的起止时间段上，年度化的(平均)简单收益率是多少?
   - 考虑一个投资: 1962 年初在 Alcoa 股票上投资一美元. 该投资在 1999 年末的价值为多少? 假定不存在交易费用.
4. 对 American Express 公司股票的月收益率重复前一题的相同分析.
5. 画出 American Express 公司股票从 1990 年 1 月到 1999 年 12 月每天的简单收益率和对数收益率的直方图. 把它们与有相同均值和标准差的正态分布相比较.

[20]

6. 可在芝加哥的联邦储备银行得到每天的外汇汇率. 数据是由纽约的联邦储备银行认证的纽约市场上的中午买入价. 考虑从 1994 年 1 月到 2001 年 2 月的加拿大元、德国马克、英国英镑、日元、法国法郎对美元的汇率. 汇率是每美元的外币数，只有英镑除外，它是以每英镑的美元数表示汇率的. 该数据也可以从文件"forex-c. dat"中得到.
   - 计算五种汇率序列的日对数收益率.
   - 计算这五种对数收益率序列的样本均值、方差、偏度、超出峰度、最小值和最大值.
   - 讨论这些汇率序列的经验特征.

参考文献

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press: New Jersey.

Cox, J. C., and Rubinstein, M. (1985), *Optoins Markets*, Prentice-Hall: Englewood Cliffs, New Jersey.

Sharpe, W. (1964), "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Snedecor, G. W., and Cochran, W. G. (1980), *Statistical Methods*, 7th edition, Iowa State University Press: Ames, Iowa.

[21]

第2章　线性时间序列分析及其应用

在这一章中，我们讨论线性时间序列分析的基本理论，引进对分析金融时间序列有用的简单的经济计量模型，并把这些模型用于资产收益率．概念的讨论是简短的，并强调与金融应用有关的应用．弄懂这里介绍的简单的时间序列模型对掌握以后各章中更复杂、更深入的金融计量模型是必要而有益的．已经有许多时间序列的教科书可参考．对线性时间序列分析的基本概念，参见 Box，Jenkins，and Reinsel(1994)的第 2 章和第 3 章，以及 Brockwell and Davis (1996)的第 1～3 章.

把资产收益率(如股票的对数收益率 r_t)看成随时间推移而形成的一族随机变量，我们就有一个时间序列 $\{r_t\}$．线性时间序列分析提供一个自然的框架，来研究这样一个序列的动态结构．线性时间序列的理论包括平稳性、动态相依性、自相关函数、建模和预测．我们所介绍的经济计量模型包括：(a)简单自回归(AR)模型，(b)简单滑动平均(MA)模型，(c)混合的自回归滑动平均(ARMA)模型，(d)季节模型，(e)带时间序列误差的回归模型，(f)刻画长相依性的分数阶差分模型．对资产收益率 r_t，试图用简单模型来刻画 r_t 与 t 时刻之前所拥有的信息之间的线性关系．这里的信息可以包括 r_t 的历史值和式(1.14)中的 $\boldsymbol{Y}$，这个 $\boldsymbol{Y}$ 是描述决定资产价格的经济环境的．这样，相关系数(correlation)在理解这些模型中起着重要作用，所研究的变量与其过去值的相关系数为线性时间序列分析的重点．这些相关系数称为前后相关系数(serial correlations)或自相关系数(autocorrelations)它们是研究平稳时间序列的基本工具.

2.1　平稳性

平稳性是时间序列分析的基础．如果对所有的 t，任意正整数 k 和任意 k 个正整数 $(t_1, \cdots, t_k)$，$(r_{t_1}, \cdots, r_{t_k})$的联合分布与$(r_{t_1+t}, \cdots, r_{t_k+t})$的联合分布是相同的，则称时间序列$\{r_t\}$为严平稳的(strictly stationary)．换句话说，严平稳性要求$(r_{t_1}, \cdots, r_{t_k})$的联合分布在时间的平移变换下不变．这是一个很强的条件，难以用经验方法验证．经常假定的是平稳性的一个较弱的形式：如果 r_t 的均值和 r_t 与 r_{t-l}的协方差是不随时间变化的，其中 l 是一个任意整数，则时间序列$\{r_t\}$是弱平稳的(weakly stationary)．更具体地说，若(a)$E(r_t)=\mu$，μ 是一个常数；(b)$\mathrm{Cov}(r_t, r_{t-l})=\gamma_l$，$\gamma_l$ 只依赖于 l，则$\{r_t\}$是弱平稳的．实际中，假定我们有 T 个观察数据点$\{r_t \mid t=1,\cdots, T\}$，弱平稳性意味着数据的时间图显示出 T 个值在一个常数水平上下以相同幅度波动.

在弱平稳性的条件中，隐含地假定了 r_t 的头两阶矩是有限的．由定义可见，若 r_t 是严平稳的，且它的头两阶矩是有限的，则 r_t 也是弱平稳的．反之一般是不成立的，但如果时间序

列 r_t 是正态分布的，则弱平稳性与严平稳性是等价的. 在这本书中，我们主要考虑弱平稳序列.

协方差 $\gamma_l=\mathrm{Cov}(r_t, r_{t-l})$ 称为 r_t 的间隔为 l 的自协方差. 它具有两个重要性质：(a) $\gamma_0=\mathrm{Var}(r_t)$ 和 (b) $\gamma_{-l}=\gamma_l$. 第二个性质成立是因为 $\mathrm{Cov}(r_t, r_{t-(-l)})=\mathrm{Cov}(r_{t-(-l)}, r_t)=\mathrm{Cov}(r_{t+l}, r_t)=\mathrm{Cov}(r_{t_1}, r_{t_1-l})$，其中 $t_1=t+l$.

在金融文献中，通常假定资产收益率序列是弱平稳的. 只要有足够多的历史收益率数据，这个假定是可以用经验方法验证的，例如，我们可以把数据分成若干子样本，然后检验它们的一致性.

2.2 相关系数和自相关函数

两个随机变量 X 和 Y 的相关系数定义为

$$\rho_{x,y}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}=\frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{E(X-\mu_x)^2E(Y-\mu_y)^2}},$$

其中 μ_x 和 μ_y 分别是 X 和 Y 的均值，并且假定方差是存在的. 这个系数度量的是 X 和 Y 线性相关的程度，可以证明 $-1\leqslant\rho_{x,y}\leqslant 1$ 和 $\rho_{x,y}=\rho_{y,x}$. 若 $\rho_{x,y}=0$，则这两个随机变量是不相关的. 另外，如果 X 和 Y 都是正态随机变量，则 $\rho_{x,y}=0$ 当且仅当 X 和 Y 是相互独立的. 当我们有样本 $\{(x_t, y_t)\}_{t=1}^{T}$ 时，相关系数可以由它对应的样本相关系数相合地估计出来：

23

$$\hat{\rho}_{x,y}=\frac{\sum_{t=1}^{T}(x_t-\bar{x})(y_t-\bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t-\bar{x})^2\sum_{t=1}^{T}(y_t-\bar{y})^2}},$$

其中 $\bar{x}=\sum_{t=1}^{T}x_t/T$, $\bar{y}=\sum_{t=1}^{T}y_t/T$ 分别是 X 和 Y 的样本均值.

自相关函数(ACF)

考虑弱平稳收益率序列 r_t，当 r_t 与它的过去值 r_{t-i} 线性相关时，可以把相关系数的概念拓广到自相关系数. r_t 与 r_{t-l} 的相关系数称为 r_t 的间隔为 l 的自相关系数，通常记为 ρ_l，在弱平稳性的假定下它只是 l 的函数. 具体地说，定义

$$\rho_l=\frac{\mathrm{Cov}(r_t,r_{t-l})}{\sqrt{\mathrm{Var}(r_t)\mathrm{Var}(r_{t-l})}}=\frac{\mathrm{Cov}(r_t,r_{t-l})}{\mathrm{Var}(r_t)}=\frac{\gamma_l}{\gamma_0}, \tag{2.1}$$

这里用到弱平稳序列的性质 $\mathrm{Var}(r_t)=\mathrm{Var}(r_{t-l})$. 由定义，我们有 $\rho_0=1$，$\rho_l=\rho_{-l}$ 和 $-1\leqslant\rho_l\leqslant 1$. 另外，一个弱平稳序列 r_t 是前后不相关的当且仅当对所有 $l>0$ 有 $\rho_l=0$.

对一个给定的收益率样本 $\{r_t\}_{t=1}^{T}$，设 $\bar{r}$ 是样本均值，即 $\bar{r}=\sum_{t=1}^{T}r_t/T$，则 r_t 的间隔为 1 的样本自相关系数为

$$\hat{\rho}_1 = \frac{\sum_{t=2}^{T}(r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}.$$

在一些一般性条件下，$\hat{\rho}_1$ 是 ρ_1 的相合估计．例如，若 $\{r_t\}$ 是独立同分布(iid)序列，且 $E(r_t^2)<\infty$，则 $\hat{\rho}_1$ 是渐近地服从均值为 0、方差为 $1/T$ 的正态分布(见 Brockwell 和 Davis(1991)中的定理 7.2.2)．在实际中这个结果可用来检验原假设 H_0：$\rho_1=0$ 对备选假设 H_a：$\rho_1\neq 0$．检验统计量就是通常的 t 比，即 $\sqrt{T}\hat{\rho}_1$，它渐近地服从标准正态分布．一般地，r_t 的间隔为 l 的样本自相关系数定义为

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}, \qquad 0 \leqslant l < T-1. \tag{2.2}$$

若 $\{r_t\}$ 是一个 iid 序列，满足 $E(r_t^2)<\infty$，则对任意固定的正整数 l，$\hat{\rho}_l$ 是渐近服从均值为 0、方差为 $1/T$ 的正态分布．更一般地，若 r_t 是一个弱平稳序列，满足 $r_t = \mu + \sum_{i=0}^{q}\psi_i a_{t-i}$，其中 $\psi_0 = 1$，$\{a_j\}$ 是高斯白噪声序列，则对 $l>q$，$\hat{\rho}_l$ 渐近于均值为 0、方差为 $\left(1+2\sum_{i=1}^{q}\rho_i^2\right)/T$ 的正态分布．在时间序列的文献中，这个结果称为 Bartlett 公式(Box，Jenkins 和 Reinsel(1994))，它被用来检验 H_0：$\rho_l=0$ 与 H_a：$\rho_l\neq 0$．关于样本自相关函数的渐近分布的更多信息，见 Fuller (1976)的第 6 章和 Brockwell and Davis(1991)的第 7 章．　[24]

对有限样本，$\hat{\rho}_l$ 是 ρ_l 的有偏估计．偏差的阶是 $1/T$，这在样本容量 T 较小的时候是不容忽视的．但在大多数金融应用中，T 是相当大的，故这个偏差并不严重．

混成检验

金融应用中常需要检验 r_t 的几个自相关系数是否同时为零．Box and Pierce(1970)提出了混成检验统计量

$$Q^*(m) = T\sum_{l=1}^{m}\hat{\rho}_l^2$$

来检验原假设 H_0：$\rho_1=\cdots=\rho_m=0$ 对备选假设 H_a：对某 $i\in\{1,\cdots,m\}$，$\rho_i\neq 0$．在 $\{r_t\}$ 为满足一定矩条件的 iid 序列的假定下，$Q^*(m)$ 渐近自由度为 m 的 χ^2 随机变量．

Ljung 和 Box(1978) 为了提高有限样本检验的功效，把统计量 $Q^*(m)$ 修改成

$$Q(m) = T(T+2)\sum_{l=1}^{m}\frac{\hat{\rho}_l^2}{T-l}. \tag{2.3}$$

在实际中，m 的选择会影响 $Q(m)$ 统计的表现，常用的 m 值有好几个．模拟研究建议取 $m\approx \ln(T)$ 会有较好的功效．

称函数 $\hat{\rho}_1$，$\hat{\rho}_2$，…为 r_t 的样本自相关函数(ACF)．这个函数在线性时间序列分析中起着重要作用．事实上，一个线性时间序列模型可完全由其 ACF 决定，线性时间序列的建模就

是用样本 ACF 来刻画数据的线性动态关系. 图 2-1 所示的是 IBM 股票从 1926 年 1 月至 1997 年 12 月的每月简单收益率和对数收益率的样本自相关函数. 这两个样本自相关函数非常接近，都显示出 IBM 股票收益率的前后相关性如果有，也很小. 两个样本 ACF 的值均在两个标准差之内，说明在 5% 水平下它们不是显著的. 此外，对简单收益率，Ljung-Box 统计量 $Q(5)=5.4$ 和 $Q(10)=14.1$ 对应的 p 值分别为 0.37 和 0.17(基于自由度为 5 和 10 的 χ^2 分布). 对于对数收益率，$Q(5)=5.8$ 和 $Q(10)=13.7$ 对应的 p 值分别为 0.33 和 0.19. 联合测试证实了 IBM 股票月收益率没有显著的前后相关性. 图 2-2 所示的是芝加哥大学 CRSP 价值加权指数的同样两个月收益率，在 5% 水平下两个收益率序列都存在某些显著的前后相关性. 对简单收益率，Ljung-Box 统计量 $Q(5)=27.8$，$Q(10)=36.0$，而对数收益率，$Q(5)=26.9$，$Q(10)=32.7$，这四个检验统计量的 p 值都小于 0.000 3，显示价值加权指数的月收益率是前后相关的. 这样，股市指数的月收益率要比个股的月收益率有更强的前后相关性.

[25]

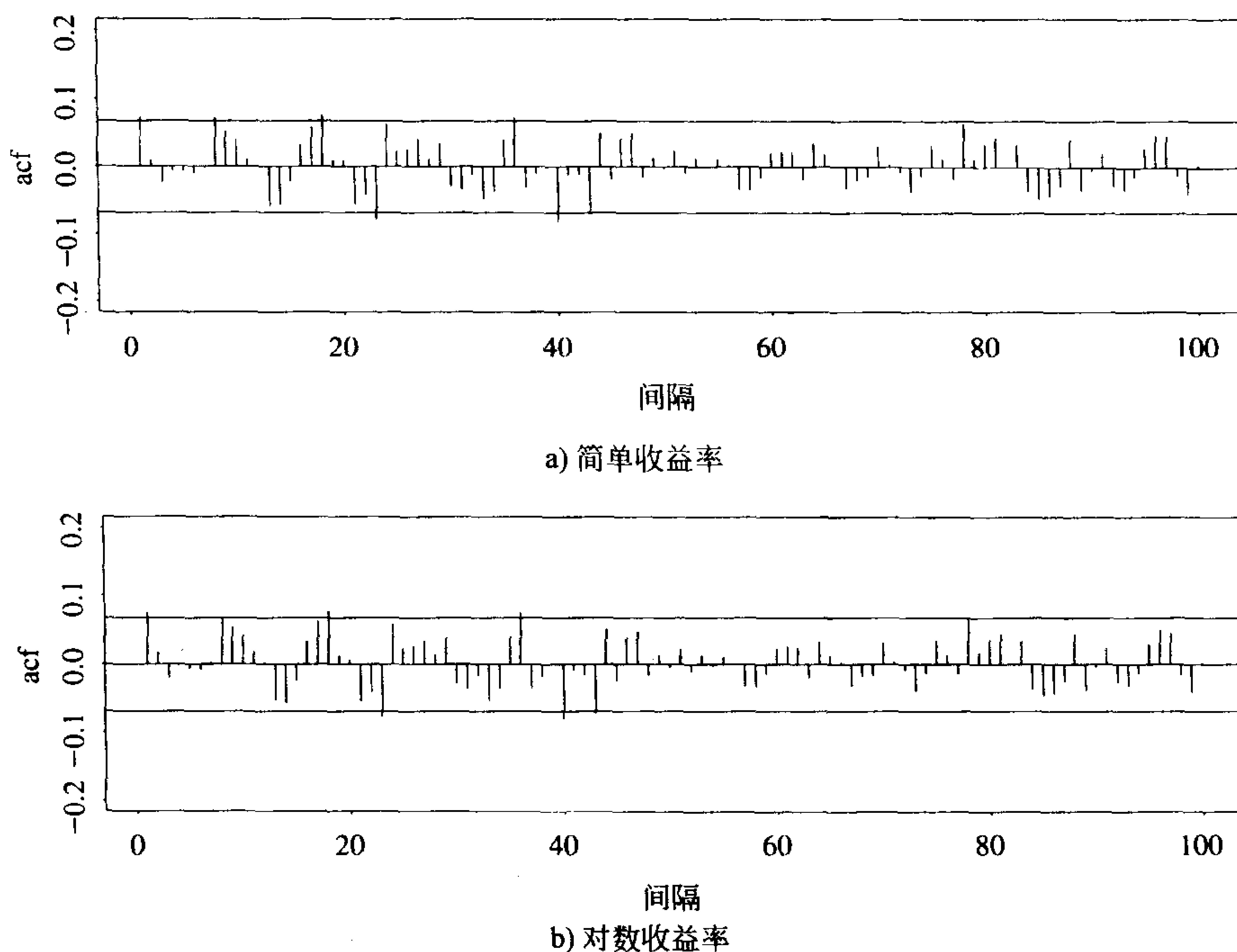


图 2-1 IBM 股票从 1926 年 1 月到 1997 年 12 月的月简单收益率和对数收益率的样本自相关函数. 每个图中的两条横线表示样本自相关函数的两个标准差的上下限

在金融文献中，资本资产定价模型(CAPM)理论就假定资产收益率序列 $\{r_t\}$ 是不可预测的，应没有自相关性. 零自相关系数的检验被用来作为有效市场假定是否成立的判定工具. 然而，股价的决定方式和指数收益率的计算方式可导致在观察到的收益率序列中有自相关性，尤其是分析高频金融数据时. 在第 5 章中我们将讨论一些这方面的问题.

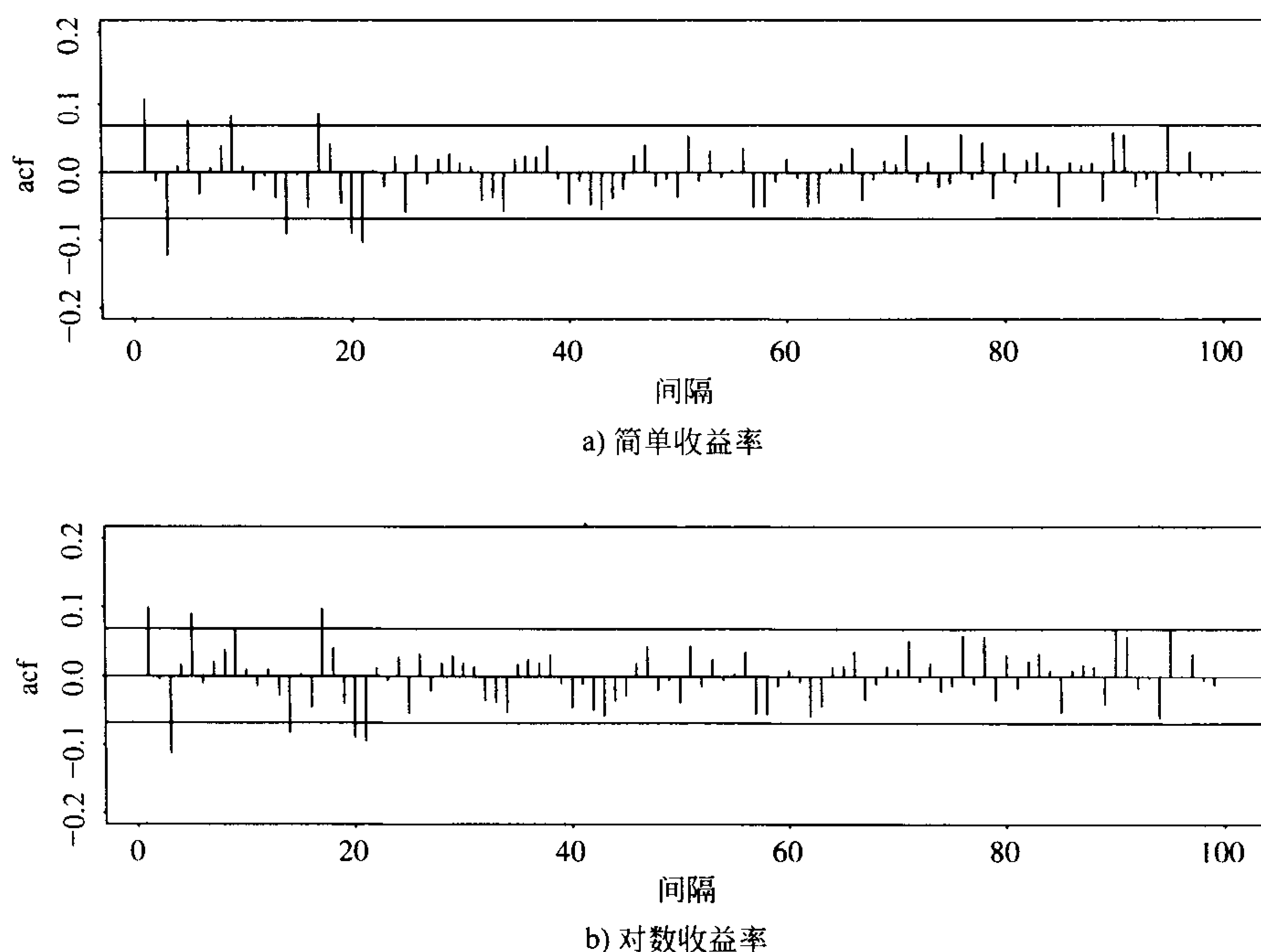


图 2-2 美国股票市场价值加权指数从 1926 年 1 月到 1997 年 12 月的月简单收益率和对数收益率的样本自相关函数. 每个图中的两条横线表示样本自相关函数的两个标准差的上下限

2.3 白噪声和线性时间序列

白噪声

如果时间序列$\{r_t\}$是一个有有限均值和有限方差的、独立同分布的随机变量序列，则称时间序列 r_t 为白噪声. 特别地，若 r_t 还服从均值为 0、方差为 σ^2 的正态分布，则称这个序列为高斯白噪声. 对白噪声序列，自相关系数为零. 在实际应用中，如果所有样本自相关函数接近于零，则认为序列是白噪声序列. 基于图 2-1 和图 2-2，IBM 股票的月收益率是接近白噪声的，而价值加权指数的收益率不是白噪声.

价值加权指数收益率的样本自相关系数的性质启示我们，在对一些资产收益率进行更深入的分析之前，有必要对其前后相关性进行建模. 下面我们讨论一些简单的时间序列模型. 它们在对时间序列的动态结构建模时是有用的，而且所述的思想在以后对资产收益率的波动率建模时也是很有用的.

线性时间序列

若时间序列$\{r_t\}$能写成

[26~27]

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \tag{2.4}$$

其中 μ 是 r_t 的均值，$\psi_0=1$，$\{a_t\}$是零均值、独立同分布的随机变量序列(即$\{a_t\}$是白噪声序列)，则称$\{r_t\}$为线性序列. 本书中我们主要关心的是 a_t 为连续型随机变量的情形. 并不是所有金融时间序列都是线性的，在第 4 章中我们研究非线性性和非线性模型.

在式(2.4)定义的线性时间序列中，系数 ψ_i 决定了 r_t 的动态结构，在时间序列的文献中这些系数称为 r_t 的 ψ 权重. 若 r_t 是弱平稳的，我们用$\{a_t\}$的独立性很容易得到 r_t 的均值和方差

$$E(r_t)=\mu, \quad \operatorname{Var}(r_t)=\sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2,$$

其中 σ_a^2 是 a_t 的方差. 另外，间隔为 l 的 r_t 的自动协方差为

$$\begin{aligned}
\gamma_l = \operatorname{Cov}(r_t, r_{t-l}) &= E\left[\left(\sum_{i=0}^{\infty} \psi_i a_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j a_{t-l-j}\right)\right] \\
&= E\left(\sum_{i,j=0}^{\infty} \psi_i \psi_j a_{t-i} a_{t-l-j}\right) \\
&= \sum_{j=0}^{\infty} \psi_{j+l} \psi_j E(a_{t-l-j}^2) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+l}.
\end{aligned}$$

因此，ψ 权重与 r_t 的自相关系数有如下关系：

$$\rho_l = \frac{\gamma_l}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+l}}{1+\sum_{i=1}^{\infty} \psi_i^2}, \qquad l \geqslant 0, \tag{2.5}$$

其中 $\psi_0=1$. 线性时间序列模型就是用来描述 r_t 的 ψ 权重的经济计量和统计模型.

2.4 简单的自回归模型

CRSP 价值加权指数的月收益率 r_t 具有统计显著的间隔为 1 的自相关系数，这个事实说明延迟的收益率 r_{t-1} 在预测 r_t 时可能会有用. 利用这样的预测功用的一个简单模型是

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t, \tag{2.6}$$

其中假设$\{a_t\}$是均值为 0、方差为 σ_a^2 的白噪声序列. 这个模型与简单的线性回归模型有相同的形式，这里 r_t 是因变量，r_{t-1} 是解释变量. 在时间序列的文献中，模型(2.6)称为简单的一阶自回归(AR)模型，或简称 AR(1)模型. 当 r_t 用它的对数波动率代替时，此简单模型也广泛地应用在随机波动率的建模中，见第 3 章和第 10 章.

[28]

式(2.6)中的 AR(1)模型有若干类似于简单线性回归模型的性质. 但是，这两个模型之间存在一些显著的差异，这一点我们将在以后讨论. 这里我们注意到：由 AR(1)模型可推得，在已知过去收益率 r_{t-1} 的条件下，

$$E(r_t \mid r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \quad \operatorname{Var}(r_t \mid r_{t-1}) = \operatorname{Var}(a_t) = \sigma_a^2.$$

也就是说，给定过去的收益率 r_{t-1}，现在的收益率将以 $\phi_0+\phi_1 r_{t-1}$ 为中心取值，离散程度以 σ_a^2

衡量. 这是 AR(1)的马尔可夫性：给定 r_{t-1}条件下，r_t 与 r_{t-i}，$i>1$ 无关. 当然，有些情况下 r_{t-1}不能单独决定 r_t 的条件期望，需要更灵活一些的模型，AR(1)模型的直接推广是 AR(p)模型：

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t, \tag{2.7}$$

其中 p 是非负整数，$\{a_t\}$如式(2.6)中的一样定义. 这个模型说的是过去的 p 个值 r_{t-i}，$i=1$，…，p 联合决定 r_t 的条件期望. AR(p)模型与以 p 个延迟值作为解释变量的多元线性回归有相同形式.

2.4.1 AR 模型的性质

为了有效地利用 AR 模型，有必要对它的基本性质进行研究. 我们详细地讨论 AR(1)和 AR(2)的性质，对一般的 AR(p)模型只给出结果.

AR(1)模型

我们首先讨论式(2.6)定义的 AR(1)模型的弱平稳性的充分必要条件. 假定序列是弱平稳的，则 $E(r_t)=\mu$，$\mathrm{Var}(r_t)=\gamma_0$，$\mathrm{Cov}(r_t, r_{t-j})=\gamma_j$，其中 μ、γ_0 是常数，γ_j 是 j 的函数而与 t 无关. 我们容易得到序列的均值、方差和自相关系数. 对式(2.6)两边取期望，因为 $E(a_t)=0$，所以

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}).$$

在平稳性的条件下，$E(r_t)=E(r_{t-1})=\mu$，从而

$$\mu = \phi_0 + \phi_1\mu \quad 或 \quad E(r_t) = \mu = \frac{\phi_0}{1-\phi_1}.$$

29

这个结果对于 r_t 有两个含义：第一，若 $\phi_1 \neq 1$，则 r_t 的均值存在；第二，r_t 的均值为 0 当且仅当 $\phi_0=0$. 这样，对平稳 AR(1)过程，常数项 ϕ_0 与 r_t 的均值有关，$\phi_0=0$ 意味着 $E(r_t)=0$.

我们利用 $\phi_0=(1-\phi_1)\mu$ 可以把 AR(1)模型写成如下形式

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + a_t. \tag{2.8}$$

重复代入，上述方程可推得

$$r_t - \mu = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots = \sum_{i=0}^{\infty} \phi_1^i a_{t-i}. \tag{2.9}$$

这样，$r_t - \mu$ 是 $a_{t-i}(i \geqslant 0)$的线性函数. 利用这个性质和序列$\{a_t\}$的独立性，我们有 $E[(r_t-\mu)a_{t+1}]=0$. 由平稳性的假定，$\mathrm{Cov}(r_{t-1}, a_t)=E[(r_{t-1}-\mu)a_t]=0$. 这个性质可以直观地看出，因为 r_{t-1}发生在 t 时刻之前，而 a_t 不依赖于过去的信息. 对等式(2.8)两边平方，然后取期望得到

$$\mathrm{Var}(r_t) = \phi_1^2 \mathrm{Var}(r_{t-1}) + \sigma_a^2,$$

其中 σ_a^2 是 a_t 的方差，这里利用 r_{t-1}与 a_t 的协方差为零的事实. 而在平稳性的假定下，$\mathrm{Var}(r_t)=\mathrm{Var}(r_{t-1})$，故

$$\mathrm{Var}(r_t) = \frac{\sigma_a^2}{1-\phi_1^2}$$

在 $\phi_1^2<1$ 时成立. 随机变量的方差是非负有限的，这就要求 $\phi_1^2<1$. 这样，AR(1)模型的弱平

稳性可推得$-1<\phi_1<1$. 反之，若$-1<\phi_1<1$，由式(2.9)和序列$\{a_t\}$的独立性，我们可以证明r_t的均值和方差是有限的，再用柯西-施瓦兹不等式，r_t的所有自协方差也是有限的. 从而，AR(1)模型是弱平稳的. 总之，式(2.6)定义的AR(1)模型弱平稳性的充分必要条件是$|\phi_1|<1$.

AR(1)模型的自相关函数

在式(2.8)两端乘以a_t，利用a_t与r_{t-1}的独立性，再取期望，我们有

30

$$E[a_t(r_t-\mu)]=\phi_1 E[a_t(r_{t-1}-\mu)]+E(a_t^2)=E(a_t^2)=\sigma_a^2,$$

其中σ_a^2是a_t的方差. 对式(2.8)两端同乘以$r_{t-l}-\mu$，取期望，再利用上述结果，我们有

$$\gamma_l=\begin{cases}\phi_1\gamma_1+\sigma_a^2, & \text{当 } l=0 \text{ 时}\\ \phi_1\gamma_{l-1}, & \text{当 } l>0 \text{ 时,}\end{cases}$$

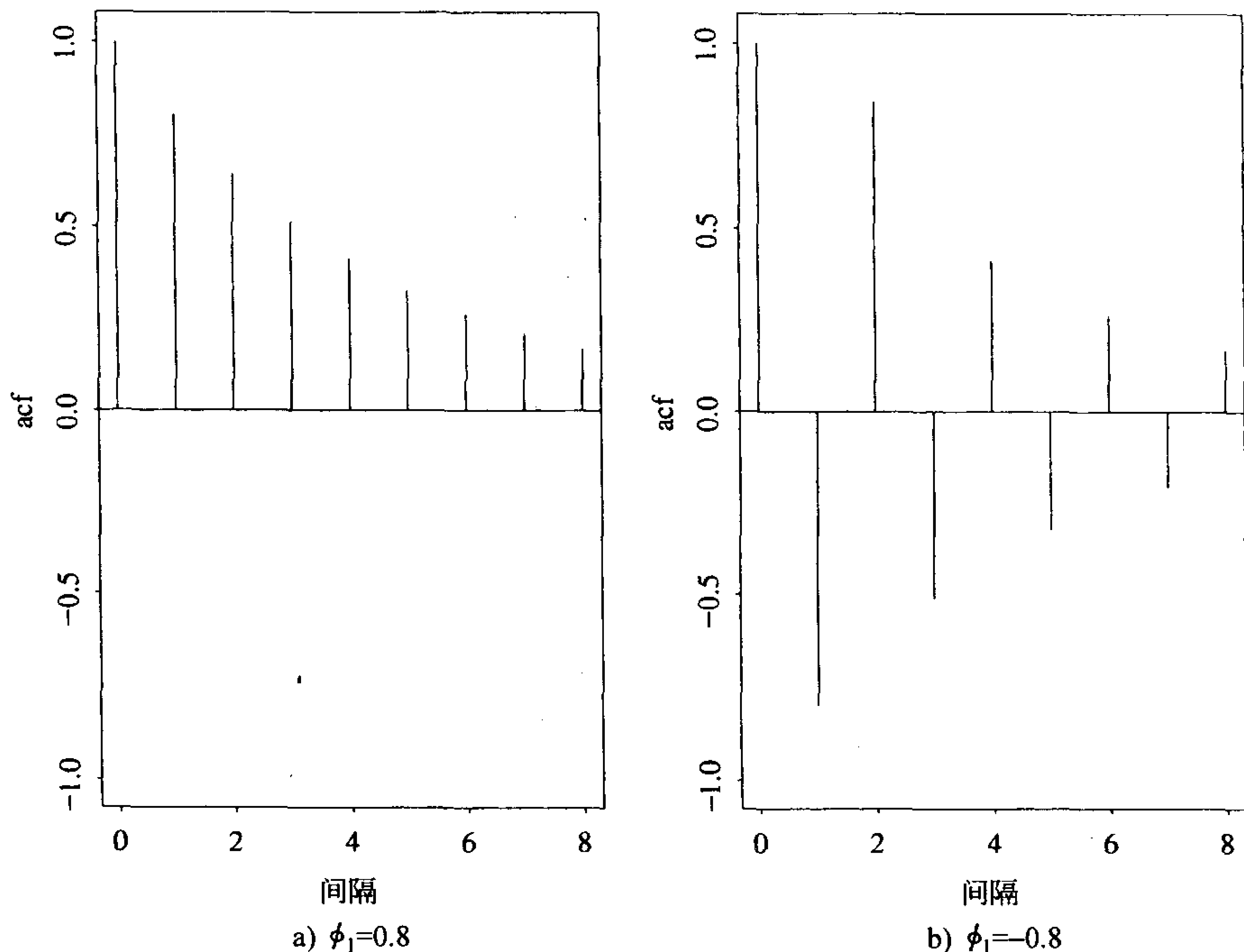
这里我们用了$\gamma_l=\gamma_{-l}$这个性质. 因此，对式(2.6)定义的弱平稳的AR(1)模型，有

$$\mathrm{Var}(r_t)=\gamma_0=\frac{\sigma^2}{1-\phi_1^2}, \quad 且\ \gamma_l=\phi_1\gamma_{l-1}, \qquad l>0.$$

由后一个方程，r_t的自相关函数(ACF)满足

$$\rho_l=\phi_1\rho_{l-1}, \qquad l>0.$$

因$\rho_0=1$，故有$\rho_l=\phi_1^l$. 这个性质表明弱平稳AR(1)序列的自相关函数从$\rho_0=1$开始以比率为ϕ_1的指数速度衰减. 对正的ϕ_1，AR(1)模型的自相关函数图像呈现漂亮的指数衰减. 对负的ϕ_1，AR(1)的ACF由上下两个都以ϕ_1^2比率衰减的图像组成. 图2-3所示的是当$\phi_1=0.8$和$\phi_1=-0.8$时的两个AR(1)模型的ACF图.



31

图 2-3 AR(1)模型的自相关函数

AR(2)模型

AR(2)模型形如

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t. \tag{2.10}$$

用与 AR(1)情形相同的方法，我们得到：只要 $\phi_1 + \phi_2 \neq 1$，有

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}.$$

利用 $\phi_0 = (1 - \phi_1 - \phi_2)\mu$，我们可把 AR(2)模型改写成

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t.$$

在上式两端同乘以 $r_{t-l} - \mu$，我们有

$$(r_{t-l} - \mu)(r_t - \mu) = \phi_1(r_{t-l} - \mu)(r_{t-1} - \mu) + \phi_2(r_{t-l} - \mu)(r_{t-2} - \mu) + (r_{t-l} - \mu)a_t.$$

再取期望，用 $E[(r_{t-l} - \mu)a_t] = 0$，$l > 0$，我们得到

$$\gamma_l = \phi_1 \gamma_{l-1} + \phi_2 \gamma_{l-2} \quad (l > 0).$$

上式称为平稳 AR(2)模型的矩方程，把上式除以 γ_0，得到 r_t 的 ACF 的性质：

$$\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2} \quad (l > 0). \tag{2.11}$$

特别地，对间隔为 1 的 ACF 满足

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} = \phi_1 + \phi_2 \rho_1.$$

因此，对平稳的 AR(2)序列 r_t，我们有 $\rho_0 = 1$，

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

$$\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2} \quad (l \geqslant 2).$$

式(2.11)的结果是：平稳 AR(2)序列的 ACF 满足二阶差分方程

$$(1 - \phi_1 B - \phi_2 B^2)\rho_l = 0,$$

32

其中 B 是向后推移算子，使 $B\rho_l = \rho_{l-1}$. 这个差分方程决定了平稳 AR(2)时间序列的 ACF 的性质，也决定了 r_t 的预报方法. 在时间序列的文献中，有时也用 L 来表示向后推移算子. 这里 L 表示延迟算子. 如，$Lr_t = r_{t-1}$，$L\psi_k = \psi_{k-1}$.

与上面的差分方程对应的，有二次多项式方程

$$x^2 - \phi_1 x - \phi_2 = 0.$$

这个方程的解是 AR(2)模型的特征根，它们是

$$x = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}.$$

用 ω_1 和 ω_2 表示两个特征根. 如果两个 ω_i 都是实值的，则模型中二次差分方程能分解成 $(1 - \omega_1 B)(1 - \omega_2 B)$，这时 AR(2)模型可以看成两个 AR(1)模型的叠加，即一个 AR(1)模型满足另一个 AR(1)模型. 因此，这时 r_t 的 ACF 是两个指数衰减的混合. 然而，如果 $\phi_1^2 + 4\phi_2 < 0$，则 ω_1 和 ω_2 都是复数(称为复共轭对)，这时 r_t 的 ACF 将呈现出减幅的正弦和余弦波的图像. 在商业和经济的应用中，复特征根是很重要的，它们会导致商业环的出现. 对经济时间序列模型来说，复值特征根是常见的. 对由式(2.10)定义的，带一对共轭复特征根的 AR(2)模型，随机环的平均长度为

$$k = \frac{360°}{\cos^{-1}\left[\phi_1/(2\sqrt{-\phi_2})\right]},$$

其中反余弦函数的值是以度来表示的.

图 2-4 所示的是四个平稳 AR(2)模型的 ACF. 图 2-4b 是 AR(2)模型的 ACF$(1-0.6B+0.4B^2)r_t=a_t$. 因为 $\phi_1^2+4\phi_2=0.36+4\times(-0.4)=-1.24<0$，这个 AR(2)模型包含两个复特征根，从而，它的 ACF 呈现出减幅的正弦和余弦波状. 其他三个 AR(2)模型都是有实特征根的，它们的 ACF 呈指数衰减.

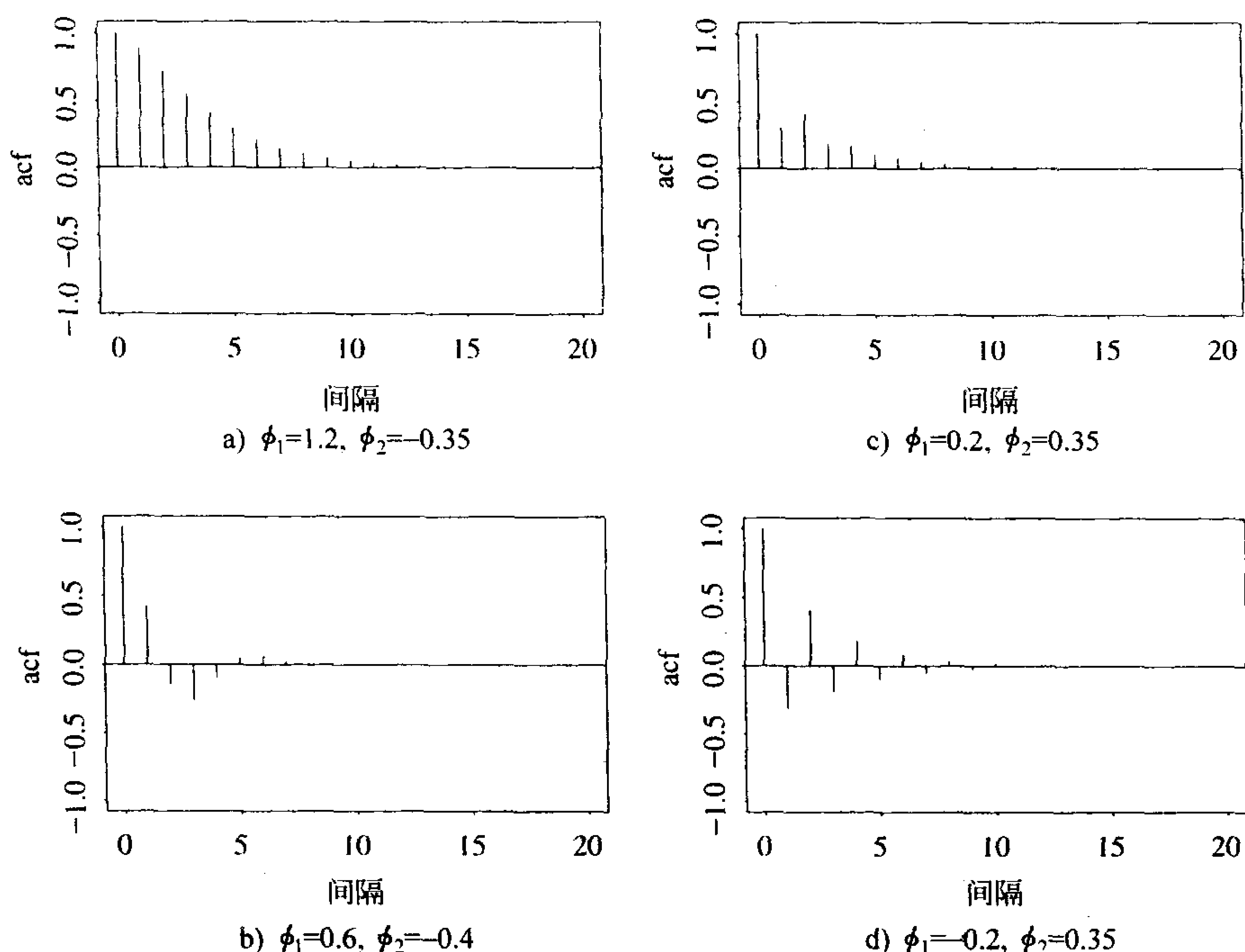


图 2-4 AR(2)模型的自相关函数

例 2.1 考虑美国的实际国民总产值(GNP)的季度增长率，时间是从 1947 年第二个季度到 1991 年的第一个季度. 这个时间序列在第 4 章中将作为一个非线性经济时间序列的例子. 这里我们简单地用 AR(3)模型来分析这组数据. 用 r_t 表示增长率，我们可以用下一小节的建立模型的方法来建立一个模型：

$$r_t = 0.0047 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0098. \tag{2.12}$$

把上述模型改写成：

$$r_t - 0.35r_{t-1} - 0.18r_{t-2} + 0.14r_{t-3} = 0.0047 + a_t,$$

我们得到对应的三阶差分方程

$$1 - 0.35B - 0.18B^2 + 0.14B^3 = 0,$$

它可以分解成

$$(1+0.52B)(1-0.87B+0.27B^2)=0.$$

第一个因子$(1+0.52B)$表示所考虑的GNP增长率大体上呈指数衰减. 对第二个因子$1-0.87B-(-0.27)B^2=0$，有$\phi_1^2+4\phi_2=0.87^2+4(-0.27)=-0.3231<0$. 因此，这个AR(3)模型的第二个因子说明美国的实际GNP的季度增长率中存在随机商业环. 这一点是合理的，因为美国经济经历了膨胀和紧缩期. 随机环的平均长度大约为

33～34

$$k=\frac{360^\circ}{\cos^{-1}\left[\phi_1/(2\sqrt{-\phi_2})\right]}=10.83(\text{季度}),$$

这大约为3年. 若我们用一个非线性模型把美国经济分解成“膨胀期”和“紧缩期”的话，数据将表明紧缩期平均长度大约为三个季度，而膨胀期的平均长度为3年(见第4章中的分析). 10.83个季度是这两个平均长度的折中. 这里得到的周期性在国民经济增长率的研究中是常见的，例如，类似上面所述的现象对OECD国家也存在.

平稳性

AR(2)时间序列平稳性的条件是它的两个特征根的模都小于1，在这个条件下，递推方程式(2.11)保证模型的自相关函数随间隔l的增加而趋于0，这个趋于0的性质是一个时间序列平稳的必要条件. 对AR(1)模型来说，这个条件也是平稳性的必要条件，这时特征多项式是$x-\phi_1=0$,特征根为$x=\phi_1$，要使r_t是平稳的，必须ϕ_1的模小于1. 而前面已证明过，对平稳AR(1)模型有$\rho_l=\phi_1^l$，当$l\to\infty$时，有$\rho_l\to 0$.

AR(p)模型

AR(1)和AR(2)模型的结果可以推广到由式(2.7)定义的AR(p)模型. 对平稳的AR(p)序列，其均值为

$$E(r_t)=\frac{\phi_0}{1-\phi_1-\cdots-\phi_p}$$

只要分母不等于0. 对应的多项式方程为

$$x^p-\phi_1x^{p-1}-\phi_2x^{p-2}-\cdots-\phi_p=0,$$

称为该AR(p)模型的特征方程. 如果这个方程的特征根都是模小于1的，则序列$\{r_t\}$是平稳的，对平稳AR(p)序列，其自相关函数满足差分方程

$$(1-\phi_1B-\phi_2B^2-\cdots-\phi_pB^p)\rho_l=0\quad(l>0).$$

自相关函数的图像呈现出正弦、余弦和指数衰减的混合状，具体形状取决于其特征根的性质.

35

2.4.2 实际中怎样识别AR模型

在实际应用中，一个AR时间序列的阶p是未知的，必须根据实际数据来决定. 这个问题叫做AR模型的阶的决定，在时间序列文献中这个问题已广泛研究过. 一般有两个决定p的方法：第一种方法是利用偏自相关函数(partial autocorrelation function)，第二种方法是用某个信息准则函数.

偏自相关函数(PACF)

时间序列的PACF是它的ACF的一个函数，在决定一个AR模型的阶p的时候，PACF是一个有用的工具. 一个简单而有效的介绍PACF的方式是考虑如下一连串的AR模型：

$$
\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t}, \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t}, \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t}, \\
r_{t_i} &= \phi_{0,4} + \phi_{1,4} r_{t-1} + \phi_{2,4} r_{t-2} + \phi_{3,4} r_{t-3} + \phi_{4,4} r_{t-4} + e_{4t}, \\
&\vdots
\end{aligned}
$$

其中，$\phi_{0,j}$是常数项，$\phi_{i,j}$是r_{t-i}的系数，e_{jt}是AR(j)模型的误差项．这些模型都是多元线性回归的形式，可用最小二乘法来估计．由于它们是按阶的大小排列的，使得我们可以应用多元线性回归分析中偏F检验的思想．第一个式子中的估计$\hat{\phi}_{11}$称为r_t的间隔为1的样本偏自相关函数；第二个式子中的估计$\hat{\phi}_{22}$称为r_t的间隔为2的样本偏自相关函数；第三个式子中的估计$\hat{\phi}_{33}$称为r_t的间隔为3的样本偏自相关函数；如此等等．

从定义可以看出，间隔为2的样本偏自相关函数$\hat{\phi}_{22}$所表示的是：在AR(1)模型$r_t=\phi_0+\phi_1 r_{t-1}+e_{1t}$基础上添加的$r_{t-2}$对$r_t$的贡献；$\hat{\phi}_{33}$表示的是在AR(2)模型上添加的$r_{t-3}$对的$r_t$贡献；如此等等．因此，对一个AR($p$)模型，间隔为$p$的样本偏自相关函数不应为零，而对所有$j>p$，$\hat{\phi}_{j,j}$应接近于零．我们利用这一性质来决定阶$p$．事实上，在一些正则性条件下，可以证明AR($p$)过程的样本偏自相关函数(PACF)有如下性质：

- 当样本容量T趋于无穷时，$\hat{\phi}_{p,p}$收敛于ϕ_p．
- 对$l>p$，$\hat{\phi}_{l,l}$收敛于零．
- 对$l>p$，$\hat{\phi}_{l,l}$的渐近方差为$1/T$．

[36] 这些结果表明AR(p)序列的样本偏自相关函数是p后截尾的．

作为例子，考虑从1926年1月到1997年12月的CRSP价值加权指数月简单收益率．表2-1给出了这个序列的样本偏自相关函数在头10个间隔上的值．$T=864$，该样本偏自相关函数的渐近标准差大约为0.03．从而，在5%的显著水平下，我们给数据识别一个AR(3)或AR(5)模型(也就是$p=3$或5)．

表2-1 从1926年1月到1997年12月的CRSP价值加权指数月简单收益率的样本偏自相关函数和AIC

p	1	2	3	4	5
PACF	0.11	−0.02	−0.12	0.04	0.07
AIC	−5.807	−5.805	−5.817	−5.816	−5.819
p	6	7	8	9	10
PACF	−0.06	0.02	0.06	0.06	−0.01
AIC	−5.821	−5.819	−5.820	−5.821	−5.818

信息准则

有几种信息准则可用来决定AR过程的阶p，它们都基于似然函数．例如，著名的Akaike信息准则(Akaike，1973)定义如下：

$$
\text{AIC} = \frac{-2}{T}\ln(\text{似然函数的最大值}) + \frac{2}{T}(\text{参数的个数}), \tag{2.13}
$$

其中 T 是样本容量. 对高斯 AR(l)模型，AIC 简化为

$$\mathrm{AIC}(l) = \ln(\hat{\sigma}_l^2) + \frac{2l}{T},$$

其中$\hat{\sigma}_l^2$ 是 σ_a^2 的最大似然估计，σ_a^2 是 a_t 的方差，T 是样本容量(见式(1.18)). 在实际应用中，事先给定一个正整数 P，对 $l=0$，1，2，…，P，计算 AIC(l)，然后选择阶 k，使 AIC 达最小值. 式(2.13)的第二项称为准则中的惩罚函数，因为它用参数的个数来惩罚所用的模型. 不同的惩罚函数导出不同的信息准则.

表 2-1 也给出了 $p=1$，2，…，10 时 AIC 的值. 这些 AIC 值都彼此很靠近，当 $p=6$ 和 $p=9$ 时，达最小值 -5.821，故按准则应选 AR(6)模型. 这个例子说明不同的方法可能会得出阶 p 的不同选择. 在实际应用中，还没有证据表明一种方法就一定比另外一种方法好. 对给定的时间序列数据选择一个 AR 模型时，还有两种因素起着重要作用，就是所研究的问题的具体信息和模型的简单性.

37

参数估计

对等式(2.7)中的一个 AR(p)模型，我们常用条件最小二乘法来估计其参数，条件最小二乘是从第 $p+1$ 个观测值开始的. 具体地说，在给定头 p 个观察值的前提下，我们有

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + a_t, t = p+1, p+2, \cdots, T,$$

其中的参数可用最小二乘法估计. 记$\hat{\phi}_i$ 为 ϕ_i 的估计，所拟合的模型为

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \cdots + \hat{\phi}_p r_{t-p},$$

对应的残差为

$$\hat{a}_t = r_t - \hat{r}_t.$$

称$\{\hat{a}_t\}$为残差序列，并得到

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{T-2p-1}.$$

作为说明，考虑给表 2-1 所示的价值加权指数的月简单收益率建立一个 AR(3)模型，所建模型为

$$r_t = 0.010\,3 + 0.104 r_{t-1} - 0.010 r_{t-2} - 0.120 r_{t-3} + \hat{a}_t, \quad \hat{\sigma}_a = 0.054.$$

系数的标准误差依次分别是 0.002，0.034，0.034 和 0.034. 除了间隔为 2 的系数外，其他参数都在 1%的水平下是显著的.

对这个例子，所建 AR 模型的系数都很小，预示着所考虑的序列的前后相关性较弱，即使这种相关性在 1%水平下是统计显著的. 所拟合的模型中$\hat{\phi}_0$ 的显著性表明这个序列的期望平均收益率是正的. 事实上，$\hat{\mu}=0.010\,3/(1-0.104+0.010+0.120)=0.01$，是一个很小的正数，但它有重要的长期意义. 这意味着所考虑的指数的长期收益率是相当大的. 利用第 1 章中定义的多周期简单收益率，这时平均年度简单毛收益率是 $\left[\prod_{t=1}^{864}(1+R_t)\right]^{12/864} - 1 \approx 0.105\,3$. 换句话说，从 1926 年到 1997 年间，CRSP 价值加权指数的月简单收益大约平均每年增长 10.53%，这支持这样一个观点：长期来看，股票市场的表现令人乐观. 1926 年初 1 美元的投资，到 1997 年底的价值大约 1 350 美元.

38

模型验证

对实际数据所拟合的模型，要仔细地验证它的合理性. 若模型是合理的，则其残差序列应是白噪声. 残差的样本自相关函数和式(2.3)定义的 Ljung-Box 统计量可用来检验$\hat{a}_t$与一个白噪声的接近程度. 对 AR(p)模型，Ljung-Box 统计量 $Q(m)$渐近服从自由度为 $m-p$ 的 χ^2 分布. 这里对自由度进行了修正，是因为有 p 个 AR 系数需要估计. 如果所拟合的模型经验证是不合理的，那么就需要对它进行改进.

现在考虑对月价值加权指数简单收益率拟合 AR(3)模型所得的残差序列. 可算得 $Q(10)=15.8$，并且基于自由度为 7 的 χ^2 分布的 p 值为 0.027. 这样，在 5%的水平下，拒绝前 10 个间隔无前后残差相关性的原假设，但在 1%的水平下，不能拒绝这个原假设. 如果模型被改进为 AR(5)模型，我们有

$$r_t = 0.0092 + 0.107r_{t-1} - 0.001r_{t-2} - 0.123r_{t-3} + 0.028r_{t-4} + 0.069r_{t-5} + \hat{a}_t,$$

而$\hat{\sigma}_a=0.054$. 间隔为 1 、3 、5 的 AR 系数在 5%的水平下是显著的，Ljung-Box 统计量 $Q(10)=11.2$,其 p 值为 0.048. 这个模型比前一模型有所改进，在 5%的水平下接近合理. 基于这个改进后的模型，r 的均值也很接近 0.01，表明所建的两个模型有相似的长期含义.

2.4.3 预测

预测是时间序列分析中的一个重要应用. 对式(2.7)中的 AR(p)模型，假定我们在时间指标为 h 的点上，欲预测 $r_{h+l}(l\geqslant 1)$. 时间指标 h 称为预测原点(forecast origin)，正整数 l 称为预测步长(forecast horizon). 设$\hat{r}_h(l)$为 r_{h+l}的最小均方误差预测，即预测函数$\hat{r}_k(l)$使得

$$E[r_{h+l}-\hat{r}_h(l)]^2 \leqslant \min_g E(r_{h+l}-g)^2,$$

其中 g 是 h 时刻(包括 t 时刻)可用信息的函数. 我们称$\hat{r}_h(l)$为 r_t 的以 h 为预测原点的向前 l 步预测.

向前 1 步预测

由 AR(p)模型，我们有

$$r_{h+1} = \phi_0 + \phi_1 r_h + \cdots + \phi_p r_{h+1-p} + a_{h+1}.$$

39

在均方损失函数下，给定模型和到 h 时刻为止的观察值，r_{h+1}的点预测即为条件期望

$$\hat{r}_h(1) = E(r_{h+1} \mid r_h, r_{h-1},\cdots) = \phi_0 + \sum_{i=1}^{p}\phi_i r_{h+1-i}$$

对应的预测误差为

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}.$$

从而，向前 1 步预测误差的方差为 $\mathrm{Var}[e_h(1)]=\mathrm{Var}(a_{h+1})=\sigma_a^2$. 若 a_t 是正态分布，则 r_{h+1}的 95%的向前 1 步区间预测是$\hat{r}_h(1)\pm 1.96\times\sigma_a$. 对式(2.4)中的线性模型，$a_{t+1}$也是以 t 为预测原点的向前 1 步预测误差. 在计量经济的文献中，a_{t+1}称为 $t+1$ 时刻序列的抖动.

在实际应用中，被估计的参数经常用来计算点预测和区间预测. 这样做的结果是条件预测，因为这样的预测没有考虑参数估计中的不确定性. 理论上，我们可以在预测中考虑参数的不确定性，但复杂性大大增加. 当在估计中使用的样本容量充分大时，条件预测与无条件预测是接近的.

向前 2 步预测

下面考虑以 h 为预测原点对 r_{h+2} 的预测．由 AR(p)模型，我们有

$$r_{h+2} = \phi_0 + \phi_1 r_{h+1} + \cdots + \phi_p r_{h+2-p} + a_{h+2}.$$

取条件期望，我们有

$$\hat{r}_h(2) = E(r_{h+2} \mid r_h, r_{h-1}, \cdots) = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + \cdots + \phi_p r_{h+2-p}$$

对应的预测误差为

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1[r_{h+1} - \hat{r}_h(1)] + a_{h+2} = a_{h+2} + \phi_1 a_{h+1}.$$

预测误差的方差为 $\mathrm{Var}[e_h(2)] = (1+\phi_1^2)\sigma_a^2$．$r_{h+2}$ 的区间预测可以用与 r_{h+1} 相同的方法计算出来．值得注意的是 $\mathrm{Var}[e_h(2)] \geqslant \mathrm{Var}[e_h(1)]$，这意味着预测步长的增加会使预测中的不确定性也增加．这与通常的感觉是一致的：对线性时间序列来说，我们在 h 时刻对 r_{h+2} 的把握不如对 r_{h+1} 的把握大．

向前多步预测

一般地，我们有

$$r_{h+l} = \phi_0 + \phi_1 r_{h+l-1} + \cdots + \phi_p r_{h+l-p} + a_{h+l}.$$

40

基于均方损失函数最小化而得到的向前 l 步预测就是给定 $\{r_{h-i}\}_{i=0}^{\infty}$ 的条件下 r_{h+l} 的条件期望，可以由下式得到

$$\hat{r}_h(l) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(l-i),$$

其中，当 $i \leqslant 0$ 时，约定 $\hat{r}_h(i) = r_{h+i}$．这个预测可用 $\hat{r}_h(i)$，$i=1, 2, \cdots, l-1$ 递推计算出来．向前 l 步预测的误差是 $e_h(l) = r_{h+l} - \hat{r}_h(l)$．可以证明：对平稳 AR($p$)模型，当 $l \to \infty$ 时，$\hat{r}_h(l)$ 收敛于 $E(r_t)$．也就是说，对这样的序列，长期的点预测趋于无条件均值．在金融文献中，这种性质称为均值逆转(mean reversion)．预测误差的方差则趋于 r_t 的无条件方差．

表 2-2 包含了对价值加权指数月简单收益率的向前 1 步至 6 步的预测结果和相应的预测误差的标准差：以 858 为预测原点，利用前 858 个观察值重新估计 AR(5)模型．实际的收益率也给出了．由于序列的弱自相关性，预测值和预测误差的标准差很快收敛到样本均值和数据的标准差．前 858 个观察值的样本均值和标准差分别为 0.009 8 和 0.055 0.

表 2-2　对 CRSP 价值加权指数月简单收益率用 AR(5)模型的向前多步预测的结果，预测原点为 858

步数	1	2	3	4	5	6
预测值	0.007 1	−0.000 8	0.008 6	0.015 4	0.014 1	0.010 0
标准误差	0.054 1	0.054 5	0.054 5	0.054 9	0.054 9	0.055 0
实际值	0.076 2	−0.036 5	0.058 0	−0.034 1	0.031 1	0.018 3

图 2-5 所示的是对价值加权指数月对数收益率的超出样本范围之外的向前 1 步至 6 步的预测值和它们的两个标准差的上下限．与表 2-2 中类似，建立下面的 AR(5)模型：

$$r_t = 0.007\,5 + 0.103 r_{t-1} + 0.002 r_{t-2} - 0.114 r_{t-3} + 0.032 r_{t-4} + 0.084 r_{t-5} + a_t$$

其中 $\sigma_a^2 = 0.054$，然后用这个模型来产生预测值．在这个例子中，预测值很接近实际值，所有实际值都在 95%的预测区间内．对这样一个具体的序列，对月简单收益率和月对数收益率的

[41] AR 模型是很接近的.

注释：前面的时间序列分析是用 SCA 软件进行的，所用命令在附录 A 中给出.

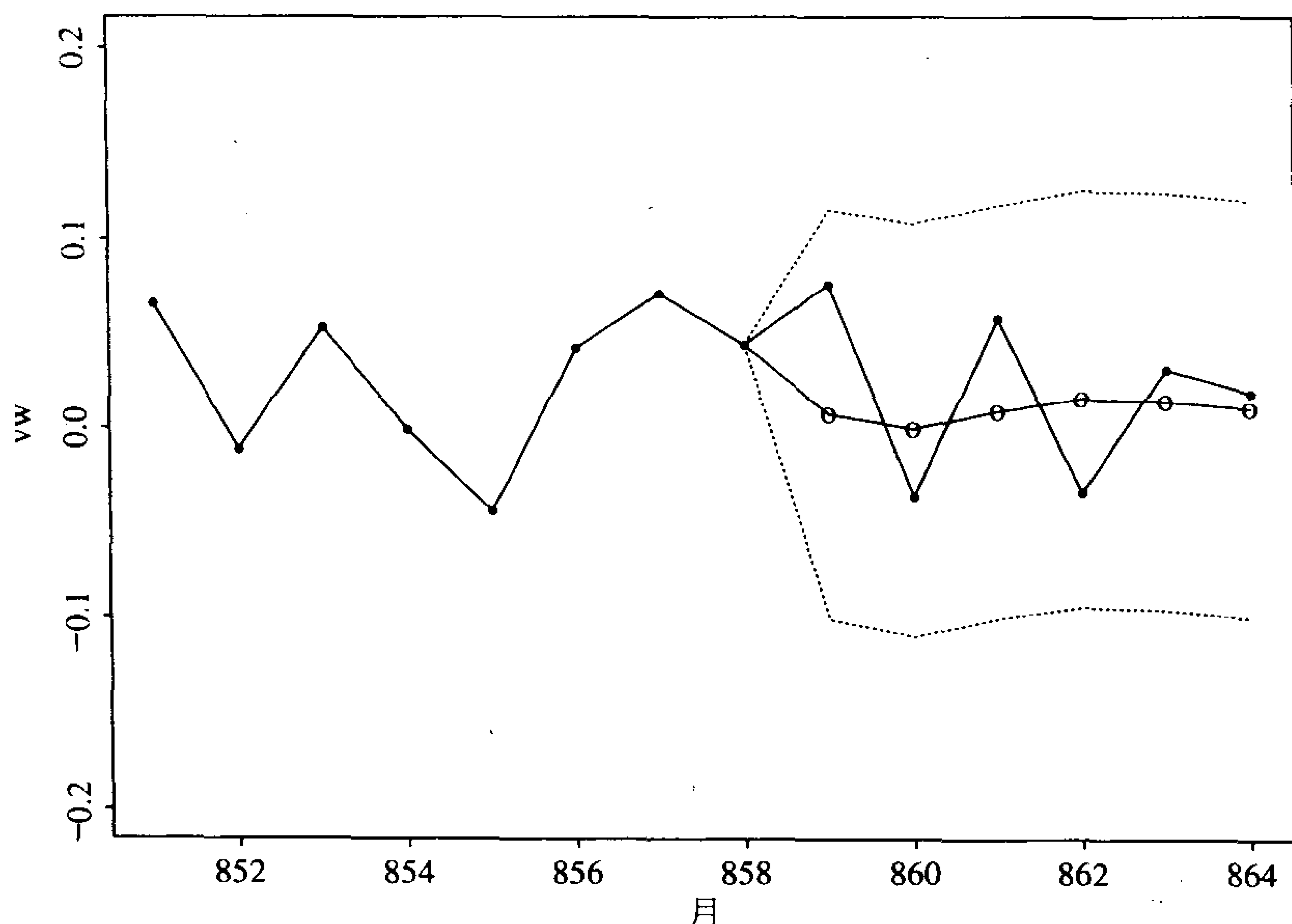


图 2-5　CRSP 价值加权指数月对数收益率的向前 1 步至 6 步的预测图. 预测原点为 $t=858$，预测值用空心点"∘"表示，实际值用实点"•"表示. 两条虚线表示预测值的两个标准误差的上下限

2.5　简单滑动平均模型

我们来讨论在金融收益率序列建模中有用的另一类简单模型. 这类模型叫做滑动平均模型(moving-average，MA). 引进 MA 模型有几种方式. 一种方式是把它当作白噪声序列的简单推广，另一种方式是把它看成参数受某种限制的无穷阶 AR 模型. 我们采用第二种方式. 如在第 5 章将要讨论的，股票交易中的讨价还价的弹跳会在收益率序列中引出一个 MA(1)结构.

除了简单性之外，没有特别的理由假定 AR 模型的阶是有限的. 我们可以设想(至少在理论上)无穷阶 AR 模型为

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + a_t.$$

然而，这样的 AR 模型是没有实际意义的，因为它有无穷多个参数. 使这样的模型有实际用途的一个方式是，假定系数 ϕ_i 满足某种限制，使得它们可由有限个参数决定. 这种想法的一个特殊情形为

[42]

$$r_t = \phi_0 - \theta_1 r_{t-1} - \theta_1^2 r_{t-2} - \theta_1^3 r_{t-3} - \cdots + a_t, \tag{2.14}$$

其中系数只依赖于单个参数 θ_1，$\phi_i = -\theta_1^i (i \geqslant 1)$. 要使式(2.14)中的模型是平稳的，$\theta_1$ 必须是绝对值小于 1 的，否则 θ_1^i 和序列本身将发散. 因为 $|\theta_1| < 1$，当 $i \to \infty$ 时，有 $\theta_1^i \to 0$，从而 r_{t-i}

对 r_t 的贡献随 i 的增加是指数型衰减的. 这一点是合理的，因为平稳序列 r_t 对它的延迟值 r_{t-i} 的依赖程度应随时间衰减.

式(2.14)中的模型能写成一个相当紧凑的形式. 为此，我们把模型先写成

$$r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \cdots = \phi_0 + a_t. \tag{2.15}$$

对 r_{t-1} 的模型为

$$r_{t-1} + \theta_1 r_{t-2} + \theta_1^2 r_{t-3} + \cdots = \phi_0 + a_{t-1}. \tag{2.16}$$

在式(2.16)两边乘以 θ_1，然后减去式(2.15)，得到

$$r_t = \phi_0(1-\theta_1) + a_t - \theta_1 a_{t-1}$$

此式说明：除去常数项外，r_t 是两个抖动 a_t 和 a_{t-1} 的加权平均. 因此，此模型称为 1 阶 MA 模型，或简称为 MA(1)模型. MA(1)模型的一般形式为

$$r_t = c_0 + a_t - \theta_1 a_{t-1} \tag{2.17}$$

其中 c_0 是一个常数，$\{a_t\}$是一个白噪声序列. 类似地，MA(2)模型的形式为

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}, \tag{2.18}$$

MA(q)模型为

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{2.19}$$

其中 $q>0$.

2.5.1 MA 模型的性质

我们只讨论简单的 MA(1)和 MA(2)模型，一般的 MA(q)模型的结果可用相同的方法得到.

平稳性

MA 模型总是弱平稳的，因为它们是白噪声序列的有限线性组合，其前两阶矩是不随时间变化的. 例如，考虑式(2.17)中的 MA(1)模型. 对这个模型两端取期望可得 $E(r_t)=c_0$，这是不随时间变化的. 在式(2.17)两端取方差，我们有 [43]

$$\mathrm{Var}(r_t) = \sigma_a^2 + \theta_1^2\sigma_a^2 = (1+\theta_1^2)\sigma_a^2.$$

这里我们用到 a_t 与 a_{t-1} 的不相关性. $\mathrm{Var}(r_t)$也是不随时间变化的. 这些讨论对一般的 MA(q)模型也适用，因此我们得到两个一般性质. 第一，MA 模型的常数项就是序列的均值[即 $E(r_t)=c_0$]；第二，MA(q)模型的方差为

$$\mathrm{Var}(r_t) = (1+\theta_1^2+\cdots+\theta_q^2)\sigma_a^2.$$

自相关函数

为简单起见，假定 MA(1)模型中 $c_0=0$. 对两端乘以 r_{t-l}，我们有

$$r_{t-l}r_t = r_{t-l}a_t - \theta_1 r_{t-l}a_{t-1}$$

再取期望，得到

$$\gamma_1 = -\theta_1\sigma_a^2 \text{ 和 } \gamma_l = 0 \text{ 对 } l>1.$$

利用上述结果和 $\mathrm{Var}(r_t)=(1+\theta_1^2)\sigma_a^2$，我们有

$$\rho_0 = 1, \quad \rho_1 = \frac{-\theta_1}{1+\theta_1^2}, \quad \rho_l = 0, \quad \text{对 } l>1.$$

这样，对 MA(1)模型，间隔为 1 的 ACF 不为 0，但所有间隔大于 1 的 ACF 都是 0. 换句话

说，MA(1)模型的 ACF 在间隔为 1 以后是截尾的. 对式(2.18)中的 MA(2)模型，自相关系数是

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}, \quad \rho_l = 0, \quad 对\ l > 2.$$

这时，在间隔为 2 时 ACF 截尾. 这个性质可推广到其他 MA 模型. 对 MA(q)模型，其 ACF 在间隔为 q 时不为 0，但对 $l>q$，$\rho_l=0$. 因此，MA(q)序列只与其头 q 个延迟值线性相关，从而它是一个“有限记忆”模型.

2.5.2 识别 MA 的阶

自相关函数是识别一个 MA 模型的阶的有用工具. 一个时间序列 r_t 具有自相关函数 ρ_l，若 $\rho_q \neq 0$ 但对 $l>q$ 有 $\rho_l=0$，则 r_t 服从一个 MA(q)模型.

[44]

图 2-6 所示的是 CRSP 相等加权指数(equal-weighted index)从 1926 年 1 月到 1997 年 12 月的月简单收益率的时间图和样本自相关函数. 在样本自相关函数图中有两条虚线，它们是两个标准误差的上、下限. 可见，该序列的自相关系数在间隔为 1、3、9 时显著. 对更大的间隔也有稍显著的，但我们在这里不考虑它们. 基于样本自相关函数，我们给该序列识别出下面的 MA(9)模型：

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_3 a_{t-3} - \theta_9 a_{t-9}.$$

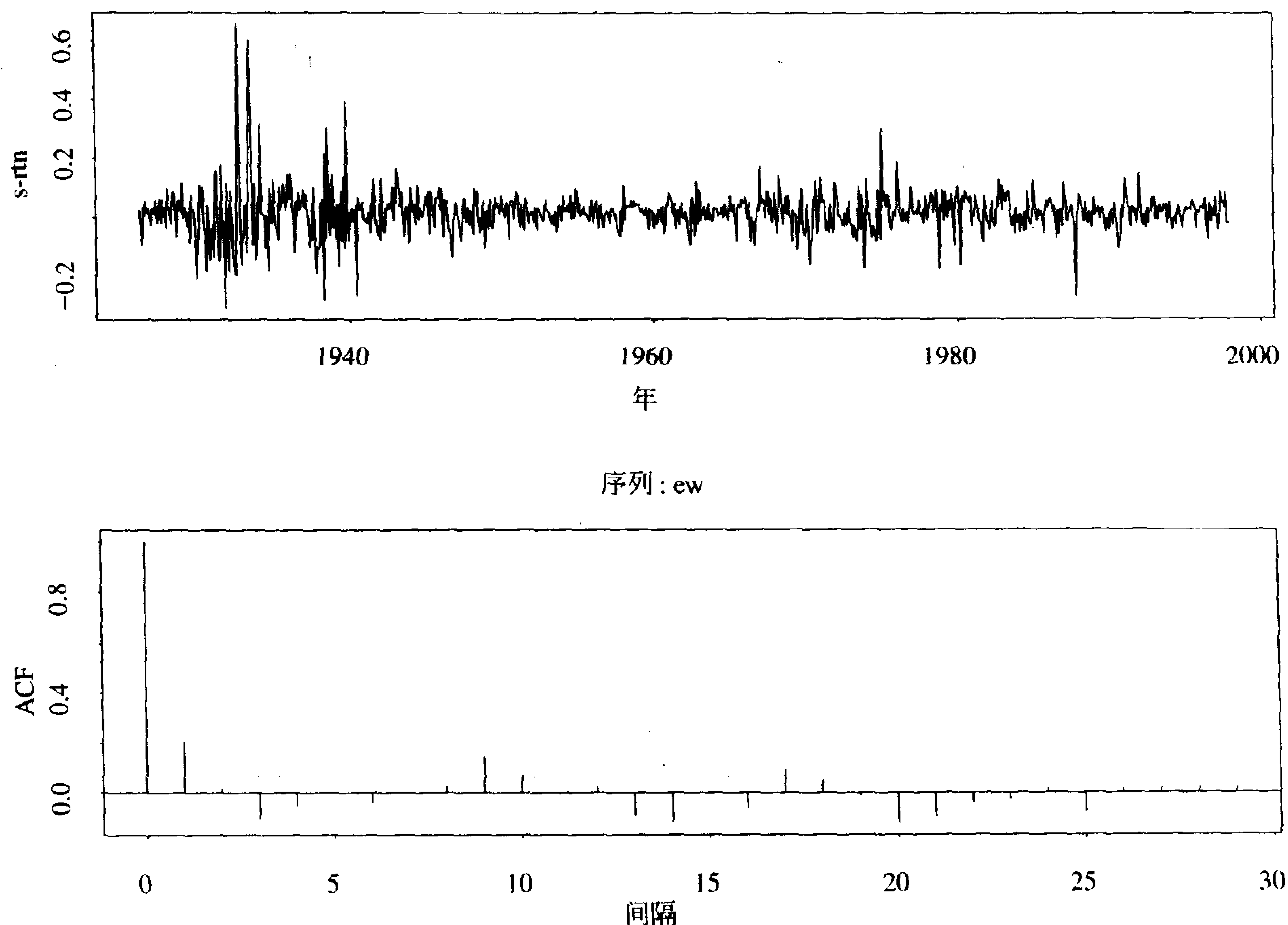


图 2-6 CRSP 相等加权指数从 1926 年 1 月到 1997 年 12 月的月简单收益率的时间图和样本自相关函数

2.5.3 估计

估计 MA 模型通常用最大似然法. 有两种方法求 MA 模型的似然函数. 第一个方法是假设初始的抖动(即 a_t, $t \leqslant 0$)都是 0, 这样 $a_1 = r_1 - c_0$, $a_2 = r_2 - c_0 + \theta_1 a_1$, …, 计算似然函数所需要的抖动可以递推得到. 这种方法称为条件似然法, 所得的估计是条件最大似然估计. 第二种方法是把初始抖动 a_t, $t \leqslant 0$ 当作模型的附加参数与其他参数一起估计出来, 这种方法称为精确似然法. 精确似然估计优于条件似然估计, 但计算会更复杂一些. 如果样本量较大, 这两种似然估计是接近的. 关于 MA 模型精确和条件似然估计的细节讨论, 读者可参阅 Box, Jenkins 和 Reinsel(1994)的书或本书第 8 章. [45]

作为例子, 我们来考虑 CRSP 相等加权指数的月简单收益率序列, 建立一个 MA(9)模型. 用条件最大似然法拟合的模型为

$$r_t = 0.0132 + a_t + 0.1775 a_{t-1} - 0.1324 a_{t-3} + 0.1349 a_{t-9}, \quad \hat{\sigma}_a = 0.0727, \quad (2.20)$$

其中系数估计的标准误差分别为 0.0030, 0.0327, 0.0328 和 0.0328. 残差的 Ljung-Box 统计量 $Q(10) = 11.4$, 基于自由度为 7 的 χ^2 分布的 p 值为 0.122. 除了有少数几个间隔(如 14, 17, 20)上残差的样本自相关函数的绝对值较大外, 模型基本上是合适的. 用精确似然法拟合的模型为

$$r_t = 0.0132 + a_t + 0.1806 a_{t-1} - 0.1315 a_{t-3} + 0.1379 a_{t-9}, \quad \hat{\sigma}_a = 0.0727 \quad (2.21)$$

其中各估计的标准误差为 0.0029, 0.0329, 0.0330 和 0.0328. 残差的 Ljung-Box 统计量为 $Q(10) = 11.6$, p 值为 0.116. 所拟合的模型也是合适的. 比较式(2.20)和式(2.21), 可见, 对这样一个具体例子来说, 条件似然法和精确似然法的差别是可以忽略的.

2.5.4 用 MA 模型预测

MA 模型的预测很容易做. 因为模型有有限记忆性, 它的点预测很快达到序列的均值. 下面来看, 设预测原点为 h, 对 MA(1)过程的向前 1 步预测, 模型为

$$r_{h+1} = c_0 + a_{h+1} - \theta_1 a_h$$

取条件期望, 我们有

$$\hat{r}_h(1) = E(r_{h+1} \mid r_h, r_{h-1}, \cdots) = c_0 - \theta_1 a_h,$$

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}.$$

[46]

向前 1 步预测误差的方差为 $\mathrm{Var}[e_h(1)] = \sigma_a^2$. 在实际中, a_h 这个量可由几个方式得到, 例如, 可假定 $a_0 = 0$, 则有 $a_1 = r_1 - c_0$, 对 $2 \leqslant t \leqslant h$, 可由 $a_t = r_t - c_0 + \theta_1 a_{t-1}$ 递推得到 a_t. 另外, 还可用 MA(1)模型的 AR 表示来计算 a_t(见 2.6.5 节).

关于向前 2 步预测, 由方程

$$r_{h+2} = c_0 + a_{h+2} - \theta_1 a_{h+1},$$

我们有

$$\hat{r}_h(2) = E(r_{h+2} \mid r_h, r_{h-1}, \cdots) = c_0,$$

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = a_{h+2} - \theta_1 a_{h+1}.$$

预测误差的方差为 $\mathrm{Var}[e_h(2)]=(1+\theta_1^2)\sigma_a^2$，这是模型的方差，它大于或等于向前 1 步预测误差的方差. 上面的结果表明 MA(1)的向前 2 步预测即是模型的无条件均值，这一点对任意预测原点 h 都正确. 一般地，$\hat{r}_h(l)=c_0(l\geqslant 2)$. 总而言之，对一个 MA(1)模型，以 h 为预测原点的向前 1 步预测为 $c_0-\theta_1 a_h$，向前多步预测为模型的无条件均值 c_0. 如果我们画出$\hat{r}_h(l)$对 l 变化的图像，会看到从一步以后预测值成一个水平直线.

类似地，对 MA(2)模型，我们有

$$r_{h+l}=c_0+a_{h+l}-\theta_1 a_{h+l-1}-\theta_2 a_{h+l-2},$$

由此得到

$$\hat{r}_h(1)=c_0-\theta_1 a_h-\theta_2 a_{h-1},$$

$$\hat{r}_h(2)=c_0-\theta_2 a_h,$$

$$\hat{r}_h(l)=c_0(l>2).$$

这样，MA(2)模型的向前两步预测即达到序列的均值，两步以后预测误差的方差也是序列的方差. 一般地，对一个 MA(q)模型，向前 q 步以后的预测就达到模型的均值.

表 2-3 给出的是式(2.20)中对相等加权指数的月简单收益率建立的 MA(9)的一些预测值，预测原点为 $h=854$. 序列的前 854 个观察值的样本均值和样本标准差分别是 0.013 1 和 0.075 7. 正如所料，该表说明：(a)向前 10 步预测值即为样本均值；(b)当预测步数增加时预测误差的标准差趋于序列的标准差.

[47]

表 2-3

步数	1	2	3	4	5
预测值	0.002 6	−0.001 6	0.023 9	0.013 3	0.007 2
标准差	0.073 0	0.074 1	0.074 1	0.074 7	0.074 7
实际值	−0.047 9	−0.021 3	0.085 1	0.044 2	0.048 6
步数	**6**	**7**	**8**	**9**	**10**
预测值	0.016 3	0.010 6	0.018 6	0.008 7	0.013 1
标准差	0.074 7	0.074 7	0.074 7	0.074 7	0.075 4
实际值	0.027 1	0.081 4	−0.025 7	−0.019 8	−0.022 6

注：相等加权指数 CRSP 的月简单收益率的 MA(9)模型的预测表现. 预测原点为 $h=854$，模型的估计用条件最大似然法得到.

小结

我们对 AR 和 MA 模型进行一个简短的总结. 我们讨论了以下一些性质：

- 对 MA 模型，ACF(自相关函数)是决定其阶的有力工具，因为对 MA(q)序列 ACF 是 q 后截尾的；
- 对 AR 模型，PACF(偏自相关函数)是决定其阶的有力工具，因为对 AR(p)过程 PACF 是 p 后截尾的；
- MA 序列总是平稳的，而对 AR 序列，当其特征根的模都小于 1 时，它是平稳的；
- 对一个平稳序列，向前多步预测收敛到序列的均值，预测误差的方差收敛于序列的方差.

2.6 简单的 ARMA 模型

在有些应用中，我们需要高阶的 AR 或 MA 模型才能充分地描述数据的动态结构. 这样就有很多参数要估计，问题变得繁琐了. 为了克服这个困难，人们提出了自回归滑动平均(ARMA)模型，见 Box，Jenkins 和 Reinsel(1994). 基本思想是把 AR 和 MA 模型的想法结合在一个紧凑的形式中，使所用参数的个数保持很少. 对金融中的收益率序列，直接用 ARMA 模型的机会较少. 然而，ARMA 模型的概念与波动率建模有密切关系. 事实上，推广的自回归条件异方差(GARCH)模型就可以认为是对序列$\{a_t^2\}$的 ARMA 模型(虽然不够标准)，第 3 章中会详细讨论这一点. 在这一节中，我们研究最简单的 ARMA(1，1)模型. 48

如果一个时间序列 r_t 满足

$$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1} \tag{2.22}$$

其中$\{a_t\}$是白噪声序列，则称 r_t 服从 ARMA(1，1)模型. 式(2.22)的左边是模型的 AR 部分，右边是 MA 部分，常数项为 ϕ_0. 为使这样一个模型有意义，要求 $\phi_1 \neq \theta_1$，否则，在方程的两端消去一个公因子，方程所决定的过程变为一个白噪声序列.

2.6.1 ARMA(1，1)模型的性质

ARMA(1，1)模型的性质是 AR(1)模型的相应性质的推广，只是作一些小的修改来处理 MA(1)部分的影响. 首先讨论平稳性条件. 在式(2.22)两端取期望，得到

$$E(r_t) - \phi_1 E(r_{t-1}) = \phi_0 + E(a_t) - \theta_1 E(a_{t-1}).$$

因为对所有的 i，$E(a_i)=0$，r_t 的均值为

$$E(r_t) = \mu = \frac{\phi_0}{1-\phi_1},$$

只要序列是弱平稳的. 这一结果和式(2.6)的 AR(1)模型的完全一样.

为简单起见，假定 $\phi_0=0$. 下面我们考虑 r_t 的自协方差函数. 首先，在模型两端乘以 a_t 再取期望，我们有

$$E(r_t a_t) = E(a_t^2) - \theta_1 E(a_t a_{t-1}) = E(a_t^2) = \sigma_a^2 \tag{2.23}$$

把模型改写成

$$r_t = \phi_1 r_{t-1} + a_t - \theta_1 a_{t-1}.$$

在上式两端取方差，得到

$$\mathrm{Var}(r_t) = \phi_1^2 \mathrm{Var}(r_{t-1}) + \sigma_a^2 + \theta_1^2 \sigma_a^2 - 2\phi_1\theta_1 E(r_{t-1}a_{t-1}),$$

这里我们用到 r_{t-1} 与 a_t 不相关这一事实. 利用等式(2.23)，我们得到

$$\mathrm{Var}(r_t) - \phi_1^2 \mathrm{Var}(r_{t-1}) = (1 - 2\phi_1\theta_1 + \theta_1^2)\sigma_a^2.$$

从而，若序列 r_t 是弱平稳的，则 $\mathrm{Var}(r_t)=\mathrm{Var}(r_{t-1})$，且

$$\mathrm{Var}(r_t) = \frac{(1-2\phi_1\theta_1+\theta_1^2)\sigma_a^2}{1-\phi_1^2}$$

49

因为方差是正的，我们需要 $\phi_1^2<1$(即$|\phi_1|<1$). 这又与 AR(1)模型的平稳性条件一样.

为了得到 r_t 的自协方差函数，我们假定 $\phi_0=0$，并在式(2.22)两端乘以 r_{t-l}，得到

$$r_t r_{t-l} - \phi_1 r_{t-1} r_{t-l} = a_t r_{t-l} - \theta_1 a_{t-1} r_{t-l}$$

对 $l=1$，在上式两端取期望，并利用 $t-1$ 时的式(2.23)，我们有

$$\gamma_1 - \phi_1 \gamma_0 = -\theta_1 \sigma_a^2$$

其中 $\gamma_l=\mathrm{Cov}(r_t, r_{t-l})$. 这个结果不同于 AR(1)情形，对 AR(1)模型有 $\gamma_1-\phi_1\gamma_0=0$. 然而，对 $l=2$，取期望后得到

$$\gamma_2 - \phi_1 \gamma_1 = 0$$

这是与 AR(1)情形一样的. 事实上，用相同的方法可得到

$$\gamma_l - \phi_1 \gamma_{l-1} = 0 \quad (l > 1) \tag{2.24}$$

用 ACF 来表达，上述结果表明：对平稳 ARMA(1, 1)模型，有

$$\rho_1 = \phi_1 - \frac{\theta_1 \sigma_a^2}{\gamma_0}, \quad \rho_l = \phi_1 \rho_{l-1} \quad (l > 1)$$

这样，ARMA(1, 1)模型的 ACF 很像 AR(1)的 ACF，不同之处仅在于它的指数衰减是从间隔 2 开始的. 因此，ARMA(1, 1)的 ACF 不能在任意有限间隔后截尾.

现在来看偏自相关函数(PACF). 可以证明：ARMA(1, 1)的 PACF 也不能在有限间隔后截尾. 它与 MA(1)的 PACF 表现很像，只是指数衰减从间隔 2 开始，而不是从间隔 1 开始.

总之，ARMA(1, 1)模型的平稳性条件与 AR(1)模型的相同，ARMA(1, 1)的 ACF 与 AR(1)的形式相似，只是这种形式是从间隔 2 处开始.

2.6.2 一般的 ARMA 模型

一般的 ARMA(p, q)模型的形式为

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i},$$

[50] 其中 $\{a_t\}$是白噪声序列，p 和 q 都是非负整数. AR 和 MA 模型是 ARMA(p, q)模型的特殊情形. 利用向后推移算子，上述模型可写成

$$(1 - \phi_1 B - \cdots - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - \cdots - \theta_q B^q) a_t, \tag{2.25}$$

模型的 AR 多项式是 $1-\phi_1 B-\cdots-\phi_p B^p$，MA 多项式是 $1-\theta_1 B-\cdots-\theta_q B^q$. 我们要求 AR 多项式和 MA 多项式没有公因子，否则模型的阶(p, q)会降低. 如 AR 模型一样，AR 多项式引进了 ARMA 模型的特征方程. 如果特征方程所有根的绝对值都小于 1，则该 ARMA 模型是弱平稳的. 这时，模型的无条件均值为 $E(r_t)=\phi_0/(1-\phi_1-\cdots-\phi_p)$.

2.6.3 识别 ARMA 模型

在决定 ARMA 模型的阶时，ACF 和 PACF 都不是很有用的. Tsay 和 Tiao(1984)提出一个新方法，利用推广的自相关函数(EACF)来决定 ARMA 过程的阶. EACF 的基本思想是相当简单的. 如果我们能得到 ARMA 模型的 AR 部分的相合估计，则能导出 MA 部分. 对所导出的 MA 序列，用 ACF 决定 MA 部分的阶.

EACF 的导出相对复杂一些，细节参见 Tsay 和 Tiao(1984). 但此函数用起来是容易的.

EACF 的结果可以用一个二维表格表示，这个表的行对应于 AR 的阶 p，列对应于 MA 的阶 q. ARMA(1, 1)模型的 EACF 的理论形式由表 2-4 给出. 这个表的主要特征是：它包含由"O"组成的三角形，并且这个三角形左上角顶点位于阶(1, 1). 我们正是用这样的特征来识别一个 ARMA 过程的阶. 一般地，对 ARMA(p, q)模型，"O"组成的三角形的左上角顶点位于(p, q)的位置.

表 2-4

	MA							
AR	0	1	2	3	4	5	6	7
0	×	×	×	×	×	×	×	×
1	×	O	O	O	O	O	O	O
2	*	×	O	O	O	O	O	O
3	*	*	×	O	O	O	O	O
4	*	*	*	×	O	O	O	O
5	*	*	*	*	×	O	O	O

注：ARMA(1, 1)模型的 EACF 理论表，其中"×"代表非零，"O"代表零，"*"代表零或者非零，它在识别阶(1, 1)时不起任何作用.

作为例子，考虑 3M 公司股票的月对数收益率从 1946 年 2 月到 1997 年 12 月的值. 共有 623 个观察值，收益率序列和它的样本自相关函数由图 2-7 所示. ACF 图显示出在 5%的水平下没有显著的前后相关性. 表 2-5 所示的是样本 EACF 的值和对应的简表. 简表是用如下记号来构造的：

51

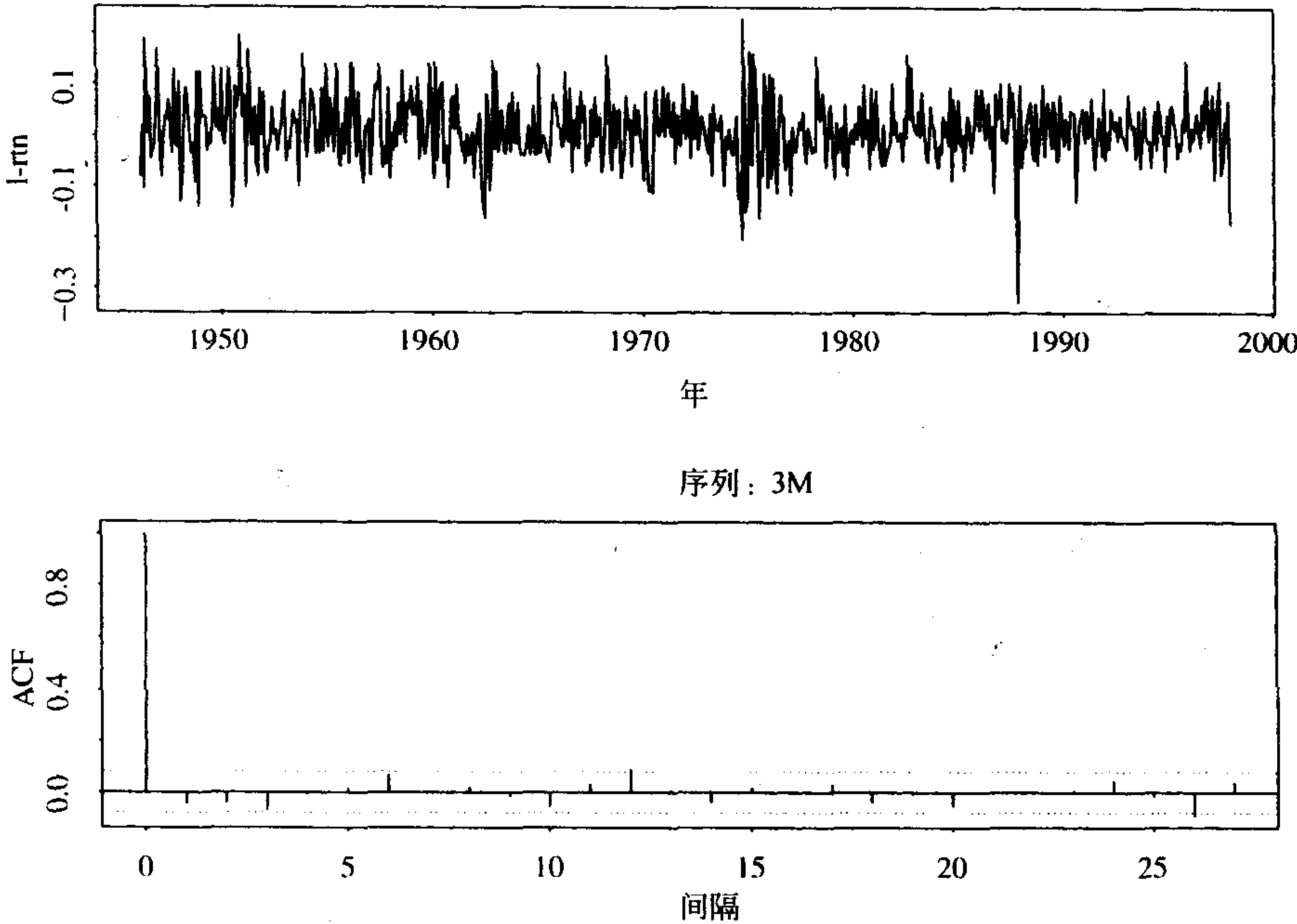


图 2-7 3M 公司股票从 1946 年 2 月到 1997 年 12 月的月对数收益率的时间图和样本自相关函数图

表 2-5 3M公司股票从1946年2月到1997年12月的月对数收益率的样本 EACF 及其简表 （MA 的阶：q）

(a)样本推广的自相关函数

p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	−0.05	−0.04	−0.07	−0.01	0.02	0.06	−0.00	0.02	−0.01	−0.06	0.03	0.09	0.01
1	−0.49	0.01	−0.06	−0.03	−0.00	0.06	0.01	0.01	−0.01	−0.05	0.02	0.08	0.02
2	−0.45	−0.18	−0.05	0.01	−0.02	0.06	0.03	0.02	−0.01	−0.00	0.01	0.05	0.05
3	−0.18	0.15	0.40	−0.01	−0.01	0.05	−0.00	0.03	−0.03	−0.00	0.00	0.02	0.05
4	0.42	0.04	0.39	−0.08	−0.01	0.01	−0.01	0.04	0.02	0.02	−0.00	0.01	0.03
5	−0.13	0.24	0.41	0.07	0.23	0.01	0.01	0.05	−0.03	0.02	−0.01	0.00	0.04
6	−0.07	−0.37	0.06	0.31	0.20	−0.09	0.01	0.06	−0.03	0.02	−0.01	0.00	0.03

(b)简化的 EACF 表

p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	O	O	O	O	O	O	O	O	O	O	O	×	O
1	×	O	O	O	O	O	O	O	O	O	O	O	O
2	×	×	O	O	O	O	O	O	O	O	O	O	O
3	×	×	×	O	O	O	O	O	O	O	O	O	O
4	×	O	×	O	O	O	O	O	O	O	O	O	O
5	×	×	×	O	×	O	O	O	O	O	O	O	O
6	O	×	O	×	×	O	O	O	O	O	O	O	O

1)“×”表示对应的 EACF 的绝对值大于或等于 $2/\sqrt{T}$，此值是 EACF 渐近标准误差的两倍；

2)“O”表示对应的 EACF 的绝对值小于 $2/\sqrt{T}$.

简表清楚地显示由“O”组成的三角形的左上角顶点位于阶$(p, q)=(0, 0)$. 只有第一行的一个“×”是例外的，它对应的的样本 EACF 值为 0.09，只比 $2/\sqrt{623}=0.08$ 稍大一点. 因此，EACF 建议给 3M 公司股票的月对数收益率服从一个 ARMA(0, 0)模型(也就是一个白噪声序列). 这与由图 2-7 中样本 ACF 所建议的是一致的.

一旦 ARMA(p, q)模型的阶确定了，它的参数就可以用条件似然法或者精确似然法来估计. 另外，残差的 Ljung-Box 统计量可以用来检验所拟合模型的合适性. 如果模型是正确的，$Q(m)$渐近地服从自由度为 $m-g$ 的 χ^2 分布，其中 g 表示模型所用的参数的个数.

2.6.4 用 ARMA 模型预测

像其 ACF 的表现一样，对 MA 部分在低步数预测的影响进行调节后，一个 ARMA(p, q)模型的预测与一个 AR(p)模型的预测有相似的特征. 设预测原点为 h，r_{h+1}的向前 1 步预测很容易由模型得到，为

$$\hat{r}_h(1)=E(r_{h+1} \mid r_h, r_{h-1}, \cdots)=\phi_0+\sum_{i=1}^{p} \phi_i r_{h+1-i}-\sum_{i=1}^{q} \theta_i a_{h+1-i}$$

相应的预测误差为 $e_h(1)=r_{h+1}-\hat{r}_h(1)=a_{h+1}$. 向前 1 步预测误差的方差为 $\mathrm{Var}[e_h(1)]=\sigma_a^2$.

对向前 l 步预测，我们有

52~53

$$\hat{r}_h(l)=E(r_{h+l}\mid r_h,r_{h-1},\cdots)=\phi_0+\sum_{i=1}^{p}\phi_i\hat{r}_h(l-i)-\sum_{i=1}^{q}\theta_i a_h(l-i)$$

其中，当 $l-i\leqslant 0$ 时，$\hat{r}_h(l-i)=r_{h+l-i}$；当 $l-i>0$ 时；$a_h(l-i)=0$；当 $l-i\leqslant 0$ 时 $a_h(l-i)=a_{h+l-i}$. 这样，ARMA 模型的向前多步预测可以递推算得．相应的预测误差为

$$e_h(l)=r_{h+l}-\hat{r}_h(l),$$

它可以用下一节给出的一个公式算出．

2.6.5 ARMA 模型的三种表示

在这一小节中，我们简单地讨论一下一个平稳 ARMA(p, q)模型的三种表示．三种表示有三种不同的用途，了解这三种表示会更好的理解 ARMA 模型．第一种表示是式(2.25)的 ARMA(p, q)模型，这个表示在参数估计时有用，在递推计算 r_t 的向前多步预测时也有用，见上一节的讨论．

对另外两种表示，我们用两个多项式比的级数展开式(长除法)．给定两个多项式 $\phi(B)=1-\sum_{i=1}^{p}\phi_iB^i$ 和$\theta(B)=1-\sum_{i=1}^{q}\theta_iB^i$，我们有(长除法)

$$\frac{\theta(B)}{\phi(B)}=1+\psi_1B+\psi_2B^2+\cdots\equiv\psi(B) \tag{2.26}$$

和

$$\frac{\phi(B)}{\theta(B)}=1-\pi_1B-\pi_2B^2-\cdots\equiv\pi(B) \tag{2.27}$$

例如，若 $\phi(B)=1-\phi_1B$ 和 $\theta(B)=1-\theta_1B$，则

$$\psi(B)=\frac{1-\theta_1B}{1-\phi_1B}=1+(\phi_1-\theta_1)B+\phi_1(\phi_1-\theta_1)B^2+\phi_1^2(\phi_1-\theta_1)B^3+\cdots$$

$$\pi(B)=\frac{1-\phi_1B}{1-\theta_1B}=1-(\phi_1-\theta_1)B-\theta_1(\phi_1-\theta_1)B^2-\theta_1^2(\phi_1-\theta_1)B^3-\cdots$$

由定义知 $\psi(B)\pi(B)=1$，利用 $Bc=c$ 对任意常数 c 成立这个事实(因为常数是不随时间变化的)，我们有

$$\frac{\phi_0}{\theta(1)}=\frac{\phi_0}{1-\theta_1-\cdots-\theta_q} \text{和} \frac{\phi_0}{\phi(1)}=\frac{\phi_0}{1-\phi_1-\cdots-\phi_p}$$

AR 表示

利用式(2.27)，ARMA(p, q)模型可写成

$$r_t=\frac{\phi_0}{1-\theta_1-\cdots-\theta_q}+\pi_1r_{t-1}+\pi_2r_{t-2}+\pi_3r_{t-3}+\cdots+a_t \tag{2.28}$$

54

这个表示说明了当前收益率 r_t 依赖于过去收益率 $r_{t-i}(i>0)$的关系．系数$\{\pi_i\}$称为 ARMA 模型的 π 权重．为了显示延迟值 r_{t-i}对 r_t 的贡献，在随 i 的增大而消失，系数 π_i 应随 i 增大而趋于零．一个 ARMA(p, q)模型如果具有这样的性质，则称为可逆的．对纯 AR 模型，$\theta(B)=1$，故 $\pi(B)=\phi(B)$，这是一个有限次的多项式，从而 $\pi_i=0(i>p)$，模型是可逆的．对其他

ARMA 模型，可逆性的充分条件是：多项式 $\theta(B)$ 的所有零点的模大于 1. 例如，对 MA(1)模型 $r_t=(1-\theta_1 B)a_t$，一次多项式 $1-\theta_1 B$ 的零点是 $B=1/\theta_1$. 从而，如果 $1/|\theta_1|>1$（即 $|\theta_1|<1$），MA(1)是可逆的.

由式(2.28)的 AR 表示，一个可逆的 ARMA 序列 r_t 是当前的扰动 a_t 与过去值的加权平均的线性组合. 对越来越远的过去值，权重呈指数衰减.

MA 表示

利用式(2.26)，ARMA(p，q)模型也能写成

$$r_t=\mu+a_t+\psi_1 a_{t-1}+\psi_2 a_{t-2}+\cdots=\mu+\psi(B)a_t, \tag{2.29}$$

其中 $\mu=E(r_t)=\phi_0/(1-\phi_1-\cdots-\phi_p)$. 这个表示具体地说明了过去的扰动 $a_{t-i}(i>0)$ 对当前值 r_t 的影响. 系数 $\{\psi_i\}$ 称为该 ARMA 模型的脉冲响应函数(impulse response function). 对弱平稳序列，系数 ψ_i 随 i 的增加呈指数衰减. 这一点是可以理解的，因为扰动 a_{t-i} 对收益率 r_t 的影响应该随时间而消失. 这样，对平稳 ARMA 模型，扰动 a_{t-i} 不能对序列有永久的影响. 如果 $\phi_0\neq 0$，这样的 MA 表示中有一个常数项，它就是 r_t 的均值（即 $\phi_0/(1-\phi_1-\cdots-\phi_p)$）.

式(2.29)的 MA 表示在计算预测误差的方差时也有用. 在预测原点 h，我们有 a_h，a_{h-1}，…. 从而，向前 l 步点预测为

$$\hat{r}_h(l)=\mu+\psi_l a_h+\psi_{l+1}a_{h-1}+\cdots, \tag{2.30}$$

相应预测误差为

$$e_h(l)=a_{h+l}+\psi_1 a_{h+l-1}+\cdots+\psi_{l-1}a_{h+1}.$$

因此，向前 l 步预测误差的方差为

$$\mathrm{Var}[e_h(l)]=(1+\psi_1^2+\cdots+\psi_{l-1}^2)\sigma_a^2, \tag{2.31}$$

[55] 它像我们所料到的一样，是预测时间长度 l 的不减函数.

式(2.29)的 MA 表示还提供了平稳时间序列均值回转的一个简单证明. 平稳性蕴涵了当 $i\to\infty$ 时，ψ_i 趋于零. 从而，由式(2.30)，我们有：当 $l\to\infty$ 时，$\hat{r}_h(l)\to\mu$. 因为 $\hat{r}_h(l)$ 是在预测原点 h 的 r_{h+l} 的条件期望，上述结果说的是，从长期来看，收益率序列预期会趋于它的均值，也就是说，序列是均值回转的. 另外，由式(2.29)的 MA 表示，我们有 $\mathrm{Var}(r_t)=\left(1+\sum_{i=1}^{\infty}\psi_i^2\right)\sigma_a^2$. 从而，由式(2.31)，当 $l\to\infty$ 时，$\mathrm{Var}[e_h(l)]\to\mathrm{Var}(r_t)$. $\hat{r}_h(l)$ 趋于 μ 的速度决定了均值回转的速度.

2.7 单位根非平稳性

到目前为止，我们的注意力还是集中在平稳的收益率序列上. 在某些研究中，利率、汇率或资产的价格序列是研究对象. 这些序列往往是非平稳的，如资产价格序列，其非平稳性主要是由于价格没有固定的水平. 在时间序列文献中，这样的非平稳序列叫做单位根(Unit-root)非平稳时间序列. 单位根非平稳时间序列最著名的例子是随机游动模型.

2.7.1 随机游动

若时间序列$\{p_t\}$满足

$$p_t = p_{t-1} + a_t, \tag{2.32}$$

其中 p_0 是一个实数，它表示这个过程的起始值，$\{a_t\}$是一个白噪声序列，则称$\{p_t\}$为一个随机游动. 若 p_t 为一支股票在第 t 天的对数价格，则 p_0 可以是该股票最初上市(initial public offering)的对数价格(简称为"对数 IPO 价格"). 若 a_t 的分布关于零点对称，则给定 p_{t-1}的条件下，上升或下降的机会各有 50%，即 p_t 将随机地上升或下降. 如果我们把随机游动模型看成一个特殊的 AR(1)模型，那么 p_{t-1}的系数是 1，这不满足 AR(1)模型弱平稳性的条件. 从而，随机游动序列不是弱平稳的，称为单位根非平稳时间序列.

随机游动模型广泛用作对数股价运动的统计模型. 在这样的模型下，股价不是可预测的或均值回转的. 下面说明：模型(2.32)在预测原点 h 的向前 1 步预测为

$$\hat{p}_h(1) = E(p_{h+1} \mid p_h, p_{h-1}, \cdots) = p_h$$

它就是预测原点的对数股价，这样的预测没有实际意义. 向前 2 步预测为

$$\begin{aligned}\hat{p}_h(2) &= E(p_{h+2} \mid p_h, p_{h-1}, \cdots) = E(p_{h+1} + a_{h+2} \mid p_h, p_{h-1}, \cdots) \\ &= E(p_{h+1} \mid p_h, p_{h-1}, \cdots) = \hat{p}_h(1) = p_h\end{aligned}$$

这又是预测原点的对数股价. 事实上，对任意的预测步长 $l>0$，都有

$$\hat{p}_h(l) = p_h.$$

这样，对所有预测步长，随机游动模型的点预测都是在预测原点序列的值. 从而，该过程不是均值回转的.

式(2.32)的随机游动模型的 MA 表示为

$$p_t = a_t + a_{t-1} + a_{t-2} + \cdots$$

这个表示有几个重要的实际含义. 首先，向前 l 步预测误差为

$$e_h(l) = a_{h+l} + \cdots + a_{h+1},$$

从而 $\mathrm{Var}[e_h(l)] = l\sigma_a^2$，当 $l\to\infty$时发散到∞. 因此，随着预测步长的增大，p_{h+l}的预测区间的长度将趋于无穷. 这表明，随着 l 的增大，点预测$\hat{p}_h(l)$变得没有用处，也再次说明该模型不是可预测的. 其次，因为当 l 增大时 $\mathrm{Var}[e_h(l)]$趋于无穷，p_t 的无条件方差是无界的. 理论上讲，这意味着对充分大的 t，p_t 可取到任何实值. 对个股的对数价格来说，这一点还说得过去，但对市场指数来讲，负的对数价格即使有也是很少见的. 从这个意义上讲，随机游动模型对市场指数的适合性值得怀疑. 第三，从上述表示看，对所有 i 有 $\psi_i=1$，这就是说，任何过去的抖动 a_{t-i}对 p_t 的影响不随时间衰减. 从而，序列有强记忆性，因为它记得所有过去的抖动. 在经济学上讲，抖动对序列有永久的效应.

2.7.2 带漂移的随机游动

如前面一些经验例子所示，市场指数的对数收益率序列会有小的正均值. 这蕴涵着对数价格的模型应为

56

$$p_t = \mu + p_{t-1} + a_t, \tag{2.33}$$

其中 $\mu=E(p_t-p_{t-1})$，$\{a_t\}$是白噪声序列．模型(2.33)中的常数项 μ 在金融研究中是很重要的．它表示的是对数价格 p_t 的时间趋势，通常称为模型的漂移(drift)．为了说明这一点，假定初始对数价格为 p_0，从而

[57]

$$\begin{aligned} p_1 &= \mu + p_0 + a_1 \\ p_2 &= \mu + p_1 + a_2 = 2\mu + p_0 + a_2 + a_1 \\ \vdots &= \quad \vdots \\ p_t &= t\mu + p_0 + a_t + a_{t-1} + \cdots + a_1 \end{aligned}$$

上式表明对数价格由时间趋势 $t\mu$ 和一个纯随机游动过程 $\sum_{i=1}^{t} a_i$ 组成．因为 $\mathrm{Var}\left(\sum_{i=1}^{t} a_i\right) = t\sigma_a^2$，其中 σ_a^2 为 a_t 的方差，p_t 的条件标准差为$\sqrt{t}\sigma_a$，它增长的速度比 p_t 的条件期望慢．从而，如果我们画出 p_t 随时间指标 t 变化的图像，我们有一个斜率为 μ 的时间趋势．正斜率 μ 蕴涵着对数价格最终趋于∞，相反，负斜率 μ 蕴涵着对数价格将随着 t 增加而趋于$-\infty$．基于这个讨论，CRSP 价值加权和相等加权指数的对数收益率序列有小的、但是统计显著的正均值这个现象就不足为奇了．

为了说明价格序列漂移参数的效应，我们考虑 3M 公司股票从 1946 年 2 月到 1997 年 12 月的月对数收益率．如表 2-5 中的样本 EACF 所示，该序列无显著的前后相关性．该序列服从简单的模型

$$r_t = 0.011\,5 + a_t, \quad \hat{\sigma}_a = 0.063\,9, \tag{2.34}$$

其中 0.011 5 是 r_t 的样本均值，其标准差为 0.002 6．3M 公司股票的月对数收益率在 1%的水平下是显著不同于零的．我们利用对数收益率序列来构造两个对数价格序列：

$$p_t = \sum_{i=1}^{t} r_i \quad 和 \quad p_t^* = \sum_{i=1}^{t} a_i,$$

其中 a_i 是式(2.34)中的均值修正对数收益率(即 $a_t=r_t-0.011\,5$)．p_t 是 3M 公司股票的对数价格，假设初始价格为 0(即 1946 年 1 月的对数价格为 0)．p_t^* 是当对数收益率的均值为 0 时，对应的对数价格．图 2-8 所示的是 p_t 和 p_t^* 的时间图，还有直线 $y_t=0.011\,5t$．从该图中可以看出，式(2.34)中常数 0.011 5 的重要性是明显的．另外，如我们所料，p_t 上升趋势的斜率大致为 0.011 5．

最后应指出的是，理解时间序列模型中常数项的意义是重要的．首先，对式(2.19)中 MA(q)模型，常数项就是序列的均值；其次，对式(2.7)中的平稳 AR(p)模型或式(2.25)中的平稳 ARMA(p，q)模型，常数项与均值有关，关系为 $\mu=\phi_0/(1-\phi_1-\cdots-\phi_p)$．另外，对带漂移的随机游动，常数项成为时间斜率．这些关于时间序列模型中常数项不同的解释，清楚地阐明了动态模型与通常的线性回归模型之间的不同．

[58]

另一个动态模型与回归模型之间的重要差别可用 AR(1)模型和简单的线性回归模型来说明：

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t, \quad y_t = \beta_0 + \beta_1 x_t + a_t$$

要使 AR(1)模型有意义，须使系数 ϕ_1 满足 $|\phi_1|\leqslant 1$，而系数 β_1 可以是任何固定的实数．

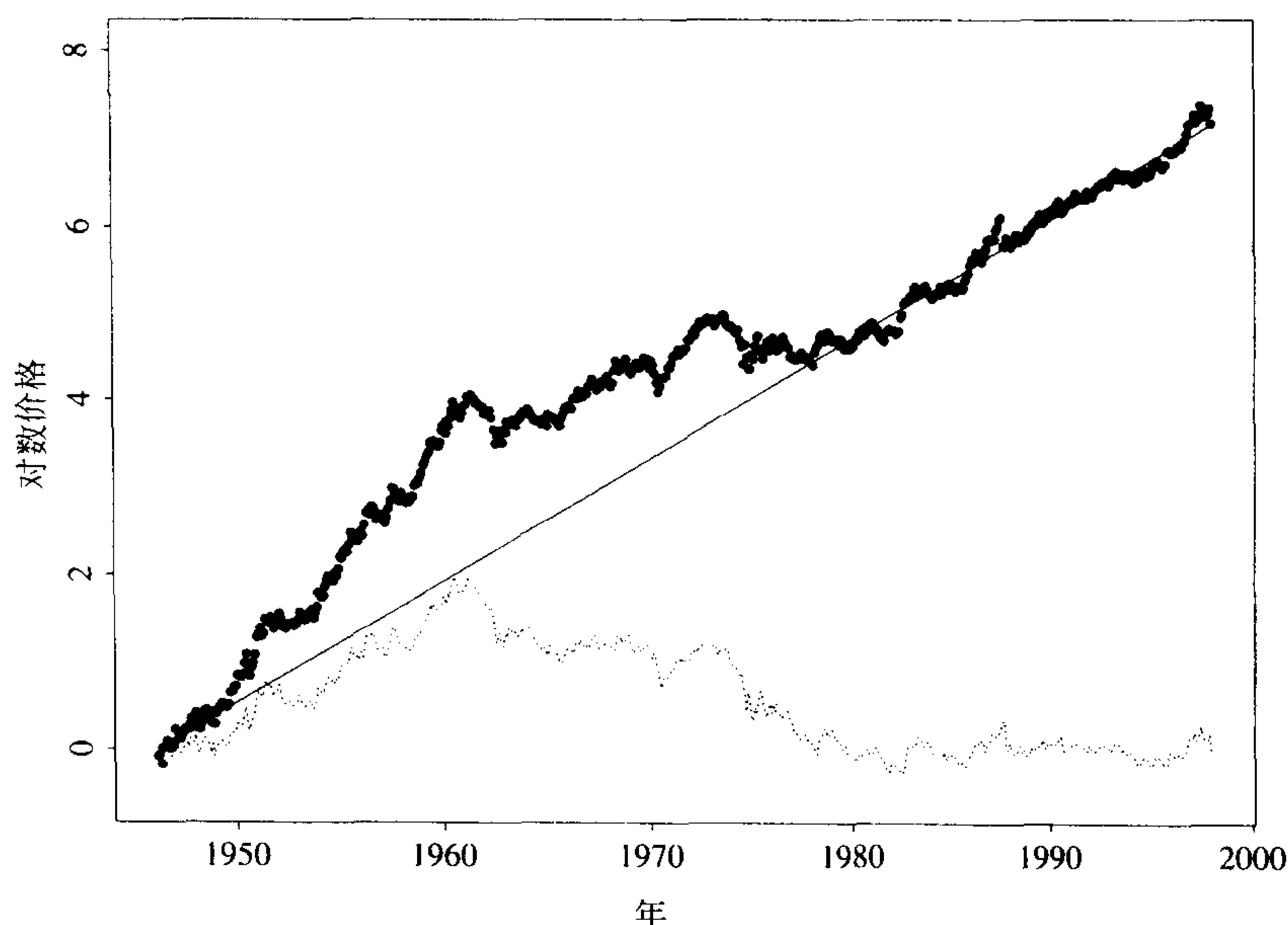


图 2-8　3M 公司股票从 1946 年 2 月到 1997 年 12 月的对数价格的时间图. 假定 1946 年 1 月的对数价格为 0. 虚线为没有时间趋势的对数价格，直线为 $y_t=0.0115t$

2.7.3　一般的单位根非平稳模型

考虑 ARMA 模型. 如果把 ARMA 模型推广到允许其 AR 多项式以 1 作为它的特征根，则模型就变成了众所周知的自回归求和滑动平均(ARIMA)模型. 因为其 AR 多项式有单位根，ARIMA 模型称为是单位根非平稳的. 像随机游动模型一样，ARIMA 模型有强记忆性，因为它的 MA 表示中的 ψ_i 系数不随时间衰减，从而模型过去的抖动 a_{t-i} 对序列有永久效应. 处理单位根非平稳性的惯用方法是用差分(differencing).

59

差分

一个时间序列 y_t 称为 ARIMA(p, 1, q)过程，如果变化序列 $c_t=y_t-y_{t-1}=(1-B)y_t$ 服从一个平稳可逆的 ARMA(p, q)模型. 在金融中，价格序列通常是非平稳的，而对数收益率序列 $r_t=\ln(p_t)-\ln(p_{t-1})$ 是平稳的. 这时，对数价格序列是单位根非平稳的，从而可当作 ARIMA 过程对待. 在时间序列文献中，通过考虑时间序列相邻两值的变化量所构成的序列，把一个非平稳序列变换成一个平稳序列，这样的思想叫做差分. 更正规地说，称 $c_t=y_t-y_{t-1}$ 为 y_t 的一阶差分序列. 在有些场合下，时间序列 y_t 可能会有多重单位根，需要做多次差分才能变成平稳序列. 例如，若 y_t 和它的一阶差分序列 $c_t=y_t-y_{t-1}$ 都是单位根非平稳的，但 $s_t=c_t-c_{t-1}=y_t-2y_{t-1}+y_{t-2}$ 是弱平稳的，则 y_t 就有双重单位根，s_t 是 y_t 的二阶差分序列. 若 s_t 服从 ARMA(p, q)模型，则 y_t 是 ARIMA(p, 2, q)过程. 对这样一个时间序列，若 s_t 有非零的均值，则 y_t 有一个二次时间函数，这二次时间函数的系数与 s_t 的均值有关. 进行季节性

调节后的美国每季度国民总产值暗含的通货紧缩(implicit price deflator)序列可能会有双重单位根，但该序列的二阶差分序列的均值不是显著地不同于零的，见本章的练习题. Box，Jenkins 和 Reinsel(1994)讨论了一般 ARIMA 模型的很多性质.

2.7.4 单位根检验

为了检验资产的对数价格 p_t 是否服从一个随机游动或一个带漂移的随机游动，我们利用如下两个模型

$$p_t = \phi_1 p_{t-1} + e_t, \tag{2.35}$$

$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t, \tag{2.36}$$

其中 e_t 为误差项，考虑原假设 H_0：$\phi_1=1$ 对备择假设 H_a：$\phi_1<1$. 这是一个著名的单位根检验问题，见 Dickey 和 Fuller(1979). 一个方便的检验统计量就是在原假设下，ϕ_1 的最小二乘估计的 t 比. 对式(2.35)，最小二乘法给出

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} p_{t-1} p_t}{\sum_{t=1}^{T} p_{t-1}^2}, \quad \hat{\sigma}_e^2 = \frac{\sum_{t=1}^{T} (p_t - \hat{\phi}_1 p_{t-1})^2}{T-1}$$

其中 $p_0=0$，T 为样本容量. t 比为

$$\mathrm{DF} \equiv t\text{ 比} = \frac{\hat{\phi}_1 - 1}{\hat{\phi}_1\text{ 的标准差}} = \frac{\sum_{t=1}^{T} p_{t-1} e_t}{\hat{\sigma}_e \sqrt{\sum_{t=1}^{T} p_{t-1}^2}},$$

60

这个 t 比通常称为 Dickey-Fuller 检验. 若 $\{e_t\}$ 是一个白噪声序列，其稍高于二阶的矩是有限的，则 DF 统计量当 $T\to\infty$ 时，趋于一个标准布朗运动的函数，更多的有关信息参见 Chan 和 Wei(1988)和 Phillips(1987). 如果 $\phi_0=0$，但采用了式(2.36)，则所得的检验 $\phi_1=1$ 的 t 比将趋于另一种非标准的渐近分布. 在上述两种情形下，都是用模拟方法来得到检验统计量的临界值，部分临界值参见 Fuller(1976)的第 8 章. 然而，如果 $\phi_0\neq 0$，且使用的是式(2.36)，则用来检验 $\phi_1=1$ 的 t 比是渐近正态的，但需要很大的样本容量来保证渐近正态分布的使用. 标准布朗运动将在本书的第 6 章中介绍.

2.8 季节模型

有些金融时间序列，如公司股票的每股每季节赢利，呈现出一定的循环或周期性. 这样的时间序列叫做季节性时间序列. 图 2-9a 所示的是 Johnson and Johnson 公司每一股份的季度赢利的时间图，时间是从 1960 年第 1 季度到 1980 年的最后一个季度. 此数据集来自于 Shumway 和 Stoffer(2000)，它有一些特殊的特征. 特别地，我们可以看到在样本的期间内赢利呈指数型上升，并有很强的季节性. 另外，随时间的推移赢利的变化性在增加. 每年重复一次循环的形状，从而此序列的周期是 4. 如果考虑月数据(如 Wal-Mart 商店的每月销售额)，则周期是 12. 季节性时间序列模型在给与天气有关的衍生产品定价和能源期货的定价方面也是有用的.

季节性时间序列的分析已有很长的历史. 在有些应用中，季节性的重要性是次要的，可把它从数据中消除，得到经季节性调整的时间序列，然后再用来做推断. 从时间序列中消除季节性的过程叫做季节调整(seasonal adjustment). 美国政府公布的多数经济数据是经季节调整的(如国民总产值的增长率和失业率). 在一些应用中，如进行预测时，季节性和数据的其他特征一样重要，必须进行合适的处理. 因为预测是金融时间序列分析的一个主要目的，故我们把注意力放在后一个话题上，讨论一些在季节性时间序列建模中有用的经济计量模型.

2.8.1 季节性差分

图 2-9b 所示的是 Johnson and Johnson 公司每股对数赢利的时间图. 我们进行对数变换有两个原因：一是处理序列的指数型增长，新的时间图确实是(取对数后)线性增长的；二是对数变换用来稳定序列的波动性，图 2-9a 中的波动在新的图中得以减弱. 对数变换在经济、金融的时间序列分析中是常用的. 在这个具体例子中，因为赢利是正的，所以在变换之前不需做调整. 而在有些场合，我们需要对每个数据点加上一个正常数后再做对数变换.

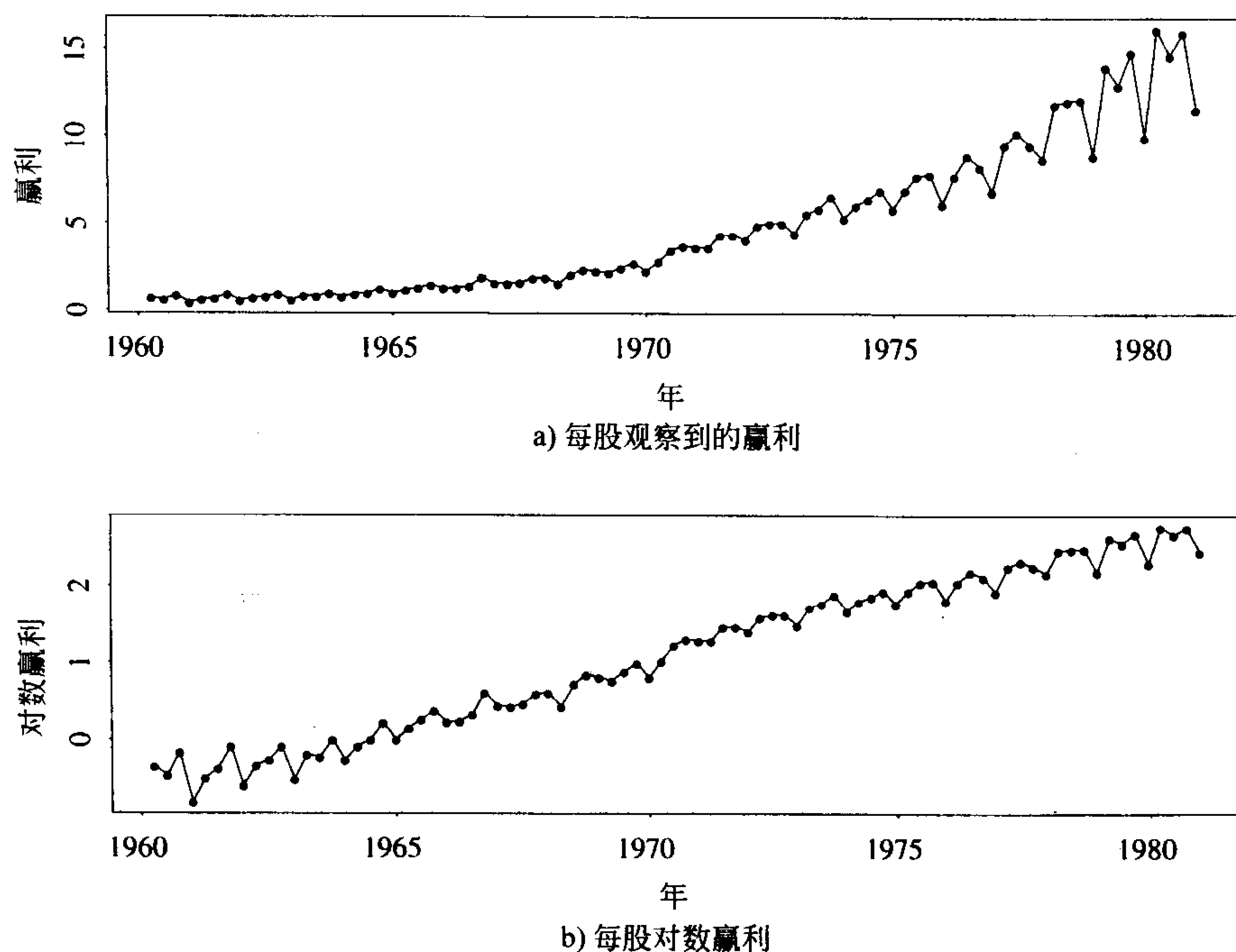


图 2-9 Johnson and Johnson 公司从 1960 年到 1980 年的每股每季节赢利的时间图

记对数赢利为 x_t. 图 2-10 的左上角是 x_t 的样本自相关函数图，它显示每股每季度对数赢利具有强的前后相关性. 处理序列的这种强序列相关性的一个惯用的方法是考虑 x_t 的一阶差分序列(即 $\Delta x_t = x_t - x_{t-1} = (1-B)x_t$). 图 2-10 的左下角是 Δx_t 的样本自相关函数图，可见当间隔是周期 4 的倍数时相关性强，这是季节性时间序列的样本自相关函数的典型表现. 按照

Box，Jenkins 和 Reinsel(1994)的第 9 章的内容，我们对 Δx_t 作另一个差分，即

$$\Delta_4(\Delta x_t) = (1-B^4)\Delta x_t = \Delta x_t - \Delta x_{t-4} = x_t - x_{t-1} - x_{t-4} + x_{t-5}$$

算子 $\Delta_4 = 1-B^4$ 叫做季节性差分(seasonal differencing)．一般地，对一个周期为 s 的季节性时间序列 y_t，季节性差分意指

61～62

$$\Delta_s y_t = y_t - y_{t-s} = (1-B^s)y_t$$

通常的差分 $\Delta y_t = y_t - y_{t-1} = (1-B)y_t$ 叫做正规差分(regular differencing)．图 2-10 的右下角是 $\Delta_4 \Delta x_t$ 的样本自相关函数，它在间隔为 1 处有一显著的负值，在间隔为 4 时有一个稍强的负相关．图 2-10 还给出了季节差分序列 $\Delta_4 x_t$ 的样本自相关函数图．

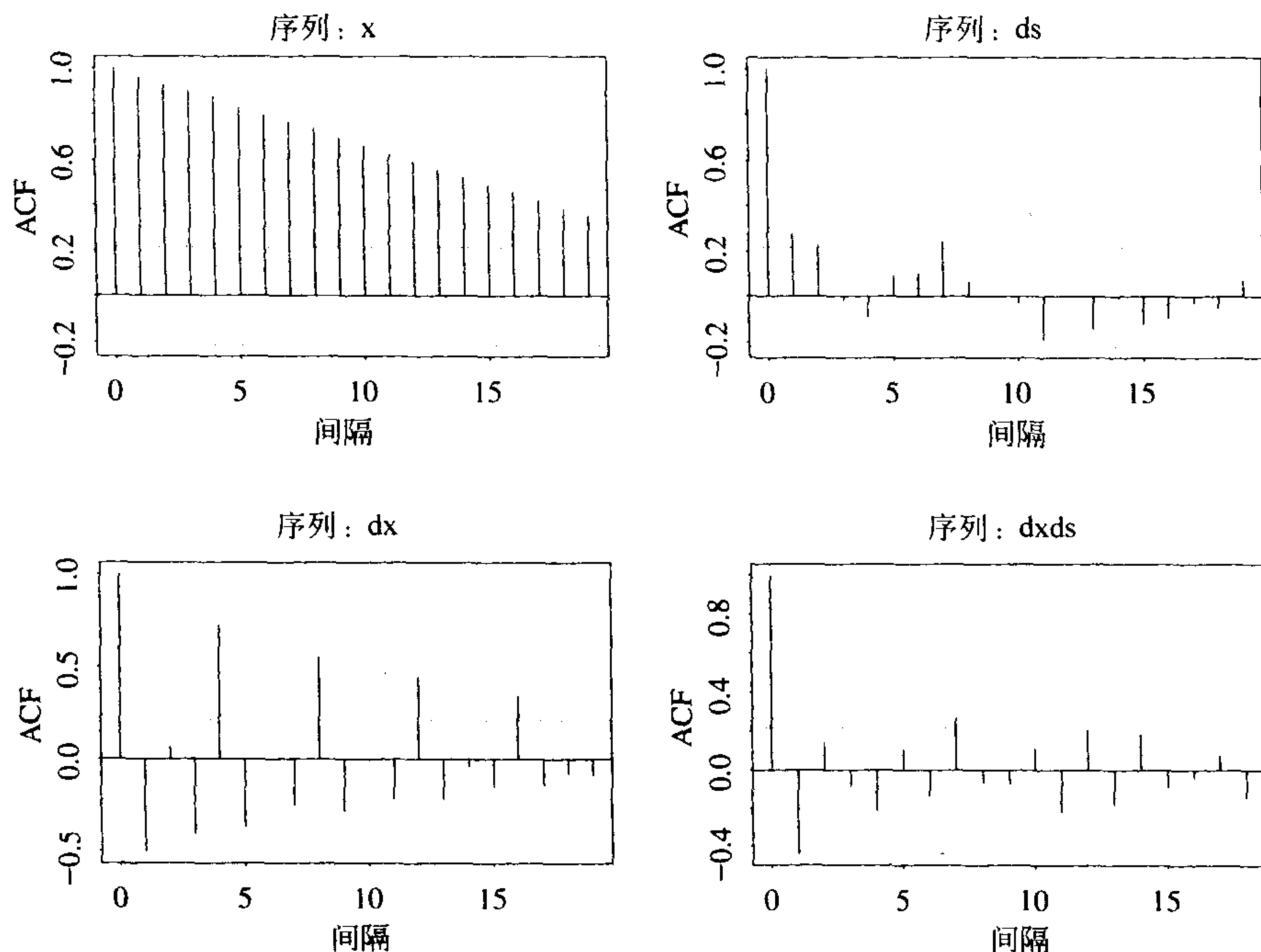


图 2-10　Johnson and Johnson 公司股票 1960 年到 1980 年每股季节赢利的对数序列的样本自相关函数，其中 x_t 是对数赢利，"dx"是一阶差分序列，"ds"是季节性差分序列，"dxds"表示经正规差分后再季节差分的序列

2.8.2　多重季节性模型

图 2-10 中 $(1-B^4)(1-B)x_t$ 的样本 ACF 的表现在季节性时间序列中是常见的．它使我们引进下面特殊的季节性时间序列模型：

$$(1-B^s)(1-B)x_t = (1-\theta B)(1-\Theta B^s)a_t \tag{2.37}$$

其中 s 是序列的周期，a_t 是白噪声序列，$|\theta|<1$，$|\Theta|<1$．此模型在文献中称为航空模型(airline model)，见 Box，Jenkins 和 Reinsel(1994)第 9 章．它在对季节性时间序列建模时被广泛地应用．此模型的 AR 部分由正规差分和季节差分两部分组成，而 MA 部分包括两个参数．

63

集中于 MA 部分，记 $w_t=(1-B^s)(1-B)x_t$，那么

$$w_t = (1-\theta B)(1-\Theta B^s)a_t = a_t-\theta a_{t-1}-\Theta a_{t-s}+\theta\Theta a_{t-s-1}$$

其中 $s>1$，容易得到 $E(w_t)=0$，并且

$$\begin{aligned}
&\text{Var}(w_t)=(1+\theta^2)(1+\Theta^2)\sigma_a^2\\
&\text{Cov}(w_t,w_{t-1})=-\theta(1+\Theta^2)\sigma_a^2\\
&\text{Cov}(w_t,w_{t-s+1})=\theta\Theta\sigma_a^2\\
&\text{Cov}(w_t,w_{t-s})=-\Theta(1+\theta^2)\sigma_a^2\\
&\text{Cov}(w_t,w_{t-s-1})=\theta\Theta\sigma_a^2\\
&\text{Cov}(w_t,w_{t-l})=0,\quad \text{对 } l\neq 0,1,s-1,s,s+1
\end{aligned}$$

因此，w_t 序列的 ACF 为

$$\rho_1=\frac{-\theta}{1+\theta^2},\quad \rho_s=\frac{-\Theta}{1+\Theta^2},\quad \rho_{s-1}=\rho_{s+1}=\rho_1\rho_s=\frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)},$$

而对 $l>0$ 但 $l\neq1$，$s-1$，s，$s+1$，有 $\rho_l=0$. 例如，如果 w_t 是季度时间序列，那么 $s=4$，且它的 ACF 只在间隔 1，3，4，5 处非零.

把上述 ACF 与 MA(1)模型 $y_t=(1-\theta B)a_t$ 和 MA(s)模型 $z_t=(1-\Theta B^s)a_t$ 的 ACF 相比较，会看出有意思的结论. y_t 和 z_t 序列的 ACF 分别是

$$\rho_1(y)=\frac{-\theta}{1+\theta^2},\quad \text{和 } \rho_l(y)=0(l>1)$$

$$\rho_s(z)=\frac{-\Theta}{1+\Theta^2},\quad \text{和 } \rho_l(z)=0(l>0,\neq s).$$

我们看到(1)$\rho_1=\rho_1(y)$，(2)$\rho_s=\rho_s(z)$，(3)$\rho_{s-1}=\rho_{s+1}=\rho_1(y)\times\rho_s(z)$. 从而，$w_t$ 在间隔 $s-1$ 和 $s+1$ 处的 ACF 可以认为是间隔为 1 和间隔为 s 的前后相关性的相互作用的结果. w_t 的模型称为多重季节性 MA 模型. 实际中，多重季节模型表示序列的正规部分与季节成分的动态是近似正交的.

模型

$$w_t=(1-\theta B-\Theta B^s)a_t,\qquad |\theta|<1,|\Theta|<1 \tag{2.38}$$

是一个非多重季节性 MA 模型. 容易看出，对式(2.38)的模型，$\rho_{s+1}=0$. 多重模型比对应的非多重模型更节省参数，因为虽然两个模型用了相同数目的参数，但多重模型的 ACF 有更多的非零点. 64

例 2.2 在此例中，我们把航空模型应用到 Johnson and Johnson 公司股票从 1960 年到 1980 年的每股季度赢利的对数序列上去. 基于精确似然法所拟合的模型为

$$(1-B)(1-B^4)x_t=(1-0.678B)(1-0.314B^4)a_t,\quad \hat{\sigma}_a=0.089,$$

其中两个 MA 参数的标准差分别为 0.080 和 0.101，残差的 Ljung-Box 统计量 $Q(12)=10.0$，其 p 值为 0.44. 模型是适合的.

为了说明上述季节性模型的预测表现，我们利用前 76 个观察值重新估计模型的参数，而把最后 8 个数据点用来进行预测评价. 以 $h=76$ 为预测原点计算向前 1 步至 8 步预测值和它们的标准误差. 利用第 1 章中给出的正态与对数正态分布之间的关系，为得到每股赢利的预测值，先进行一个反对数变换. 图 2-11 显示了模型的预测表现，其中的实线表示实际观察值，

点预测值用一些点表示，两条虚线所示的是 95%的区间预测．预测值显示出强烈的季节性，并与实际观察值相近．

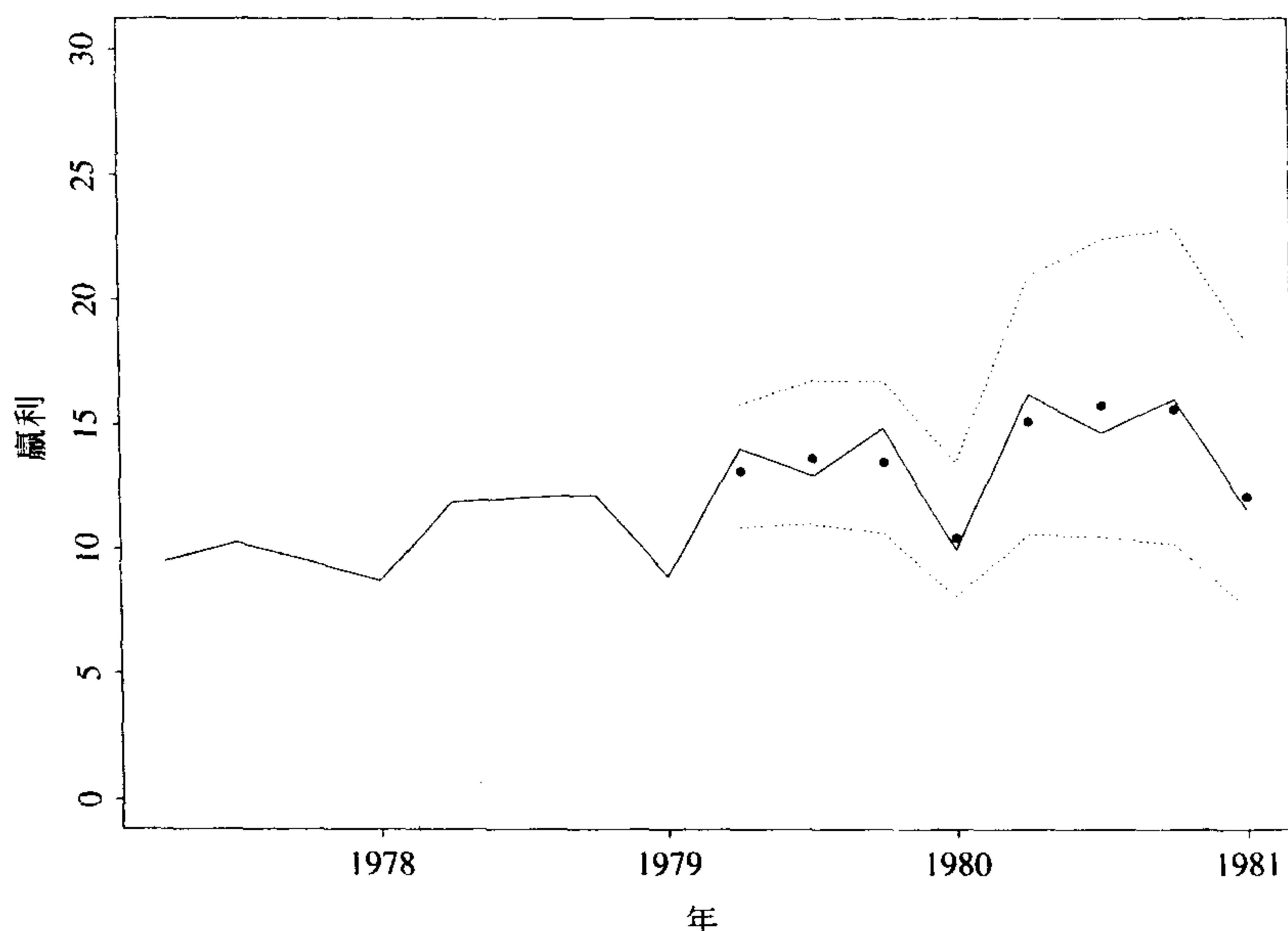


图 2-11 Johnson and Johnson 公司每股季节赢利的点预测和区间预测．预测原点是 1978 年的第四季度．实线表示实际观测值，黑点表示点预测值，两条虚线表示的是 95%的区间预测

当一个时间序列的季节性表现是随时间稳定的(即，近似于确定性函数)时候，可用哑变量(dummy variable)来处理季节性．有些分析师就采用这种方法．然而，确定的季节性只是前面讨论的多重季节性模型的一个特殊情形，如 $\Theta=1$ 时，模型(2.37)包含一个确定的季节性成分．因此，当季节性是确定的时候，用哑变量或者用多重季节性模型都能得相同的预测结果．但是，当季节性不是确定的时候，哑变量方法会导致较差的预测．在实际中，我们建议用精确似然法去估计一个多重季节性模型，特别是当样本容量较小或可能存在一个确定的季节性成分时．

2.9 带时间序列误差的回归模型

在许多应用中，主要兴趣在于两个时间序列的关系上．金融中的市场模型就是一例，它要找出个股的收益率与市场指数的收益率之间的关系．利率的期限结构是另一个例子，这时要研究的是不同期限的利率之间的关系是怎样随时间演变的．这些例子导致我们需考虑如下形式的线性回归：

$$r_{1t}=\alpha+\beta r_{2t}+e_t, \tag{2.39}$$

其中 r_{1t}和 r_{2t} 是两个时间序列，e_t 表示误差项．最小二乘(*LS*)方法经常被用来估计模型(2.39)，若$\{e_t\}$是白噪声序列，则 *LS* 方法产生相合估计．然而，在实际中经常遇到误差项 e_t 是前后相关的．这时，式(2.39)就是一个带时间序列误差的回归模型，最小二乘法(*LS*)所产

生的 α 和 β 的估计可能不是相合的.

带时间序列误差的回归模型在经济、金融中广泛应用，但它也是最经常被误用的经济计量模型之一，因为 e_t 的前后相关性经常被忽视. 应该对此模型进行仔细研究.

我们通过考虑如下两个美国的周利率序列之间的关系来介绍该模型：

1. r_{1t}：一年期固定期限国库券利率，

2. r_{3t}：三年期固定期限国库券利率.

这两个序列都是以百分数给出，都是从 1962 年 1 月 5 日至 1999 年 9 月 10 日，都有 1967 个观察值，是从圣・路易斯联邦储备银行得到的. 图 2-12 是上述两个利率序列的时间图，实线是一年期的利率，虚线是三年期的利率. 图 2-13a 是 r_{1t}对 r_{3t}的图，正好与预期的一样，这两种利率是高度相关的. 描述这两种利率之间的关系的一个自然方式是利用简单的模型 $r_{3t}=\alpha+\beta r_{1t}+e_t$，所拟合的模型是 66

$$r_{3t}=0.911+0.924r_{1t}+e_t,\quad \hat{\sigma}_e=0.538 \tag{2.40}$$

其中两个系数的标准差分别为 0.032 和 0.004，此模型的 $R^2=95.8\%$. 模型(2.40)证实了两种利率之间的高度相关性. 然而，由图 2-14 所示的该模型的残差的 ACF 及其时间图可见，该模型是严重不合适的. 残差的样本 ACF 是高度显著的，并且缓慢衰减，显示出单位根非平稳时间序列的态势. 残差的表现说明两种利率间存在明显的差异. 利用现代经济计量方法，如果我们假定两个利率序列都是单位根非平稳的，那么式(2.40)中残差的表现说明这两种利率序列不是协整的(co-integrated)(见第 8 章中关于协整的讨论). 换句话说，数据不支持关于两种利率间存在长期均衡关系的假设. 从某种意义上讲，这不出乎意料，因为“逆收益曲线”(inverted yield curve)在数据所在的时间段内出现了. 逆收益曲线指的是利率与它们离到期日的时间长度逆相关.

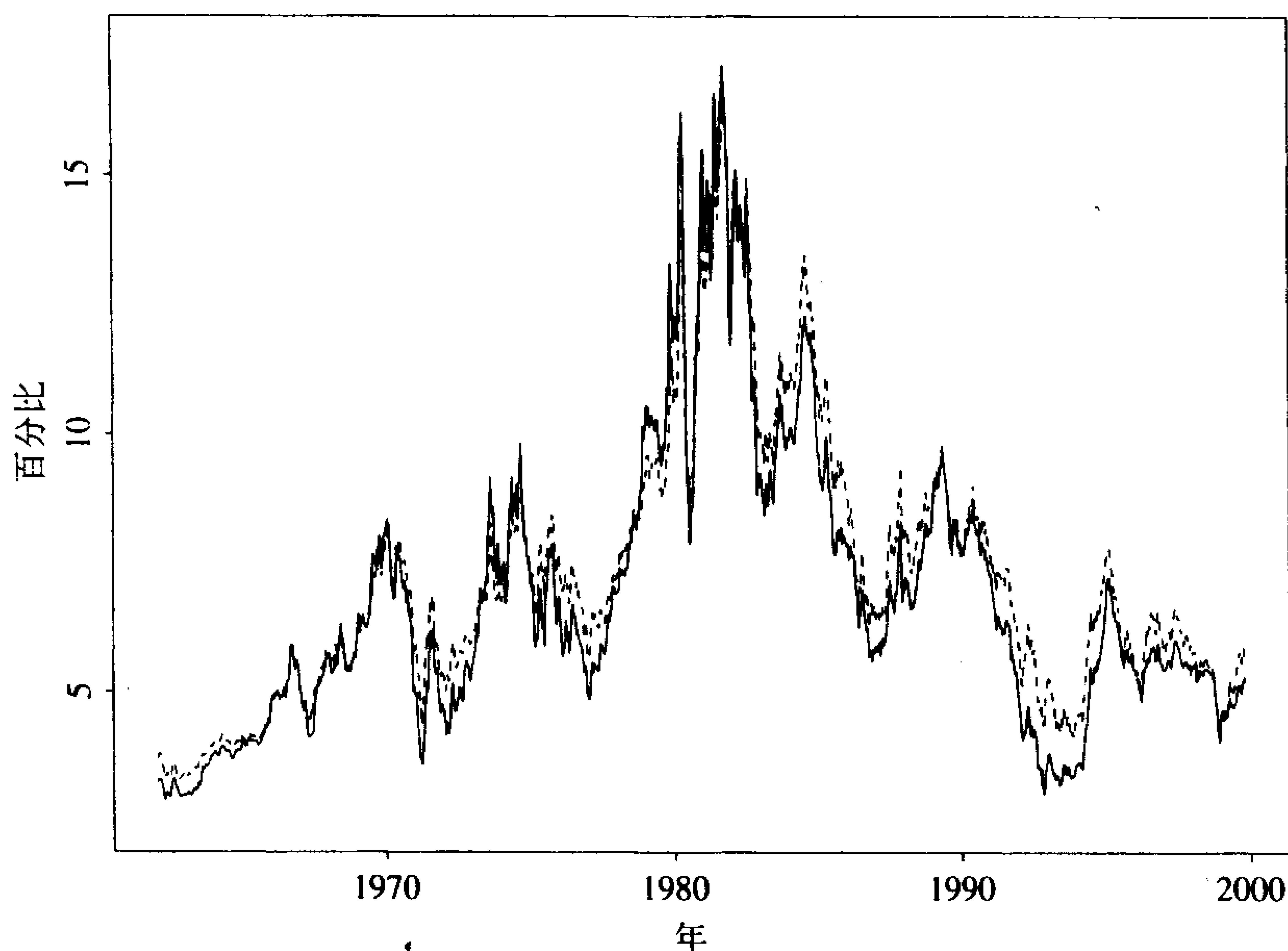


图 2-12　从 1962 年 1 月 5 日到 1999 年 9 月 10 日美国周利率(百分比)的时间图. 实线是 1 年期固定期限国库券利率，虚线是 3 年期固定期限国库券利率

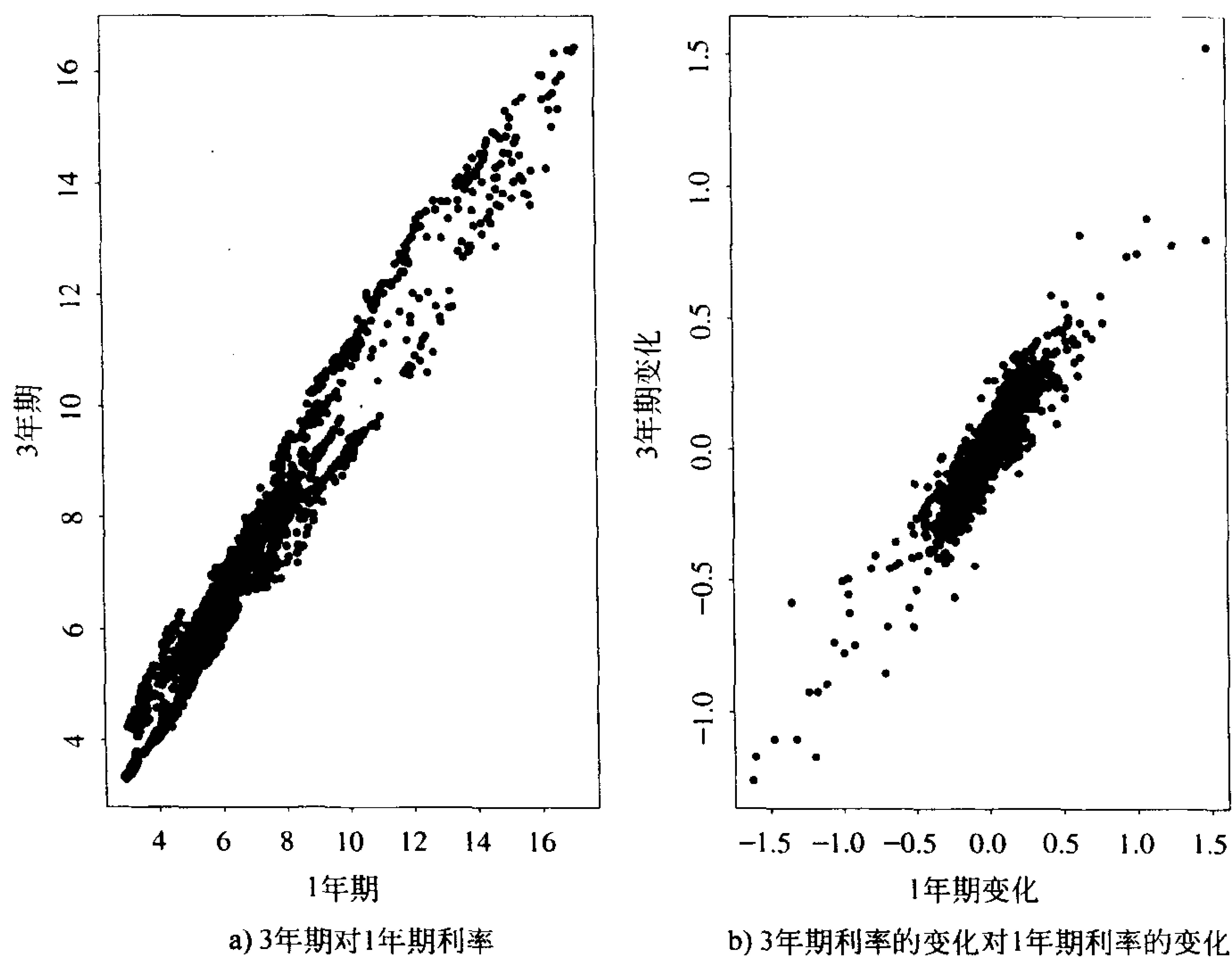


图 2-13 从 1962 年 1 月 5 日到 1999 年 9 月 10 日美国周利率的散点图

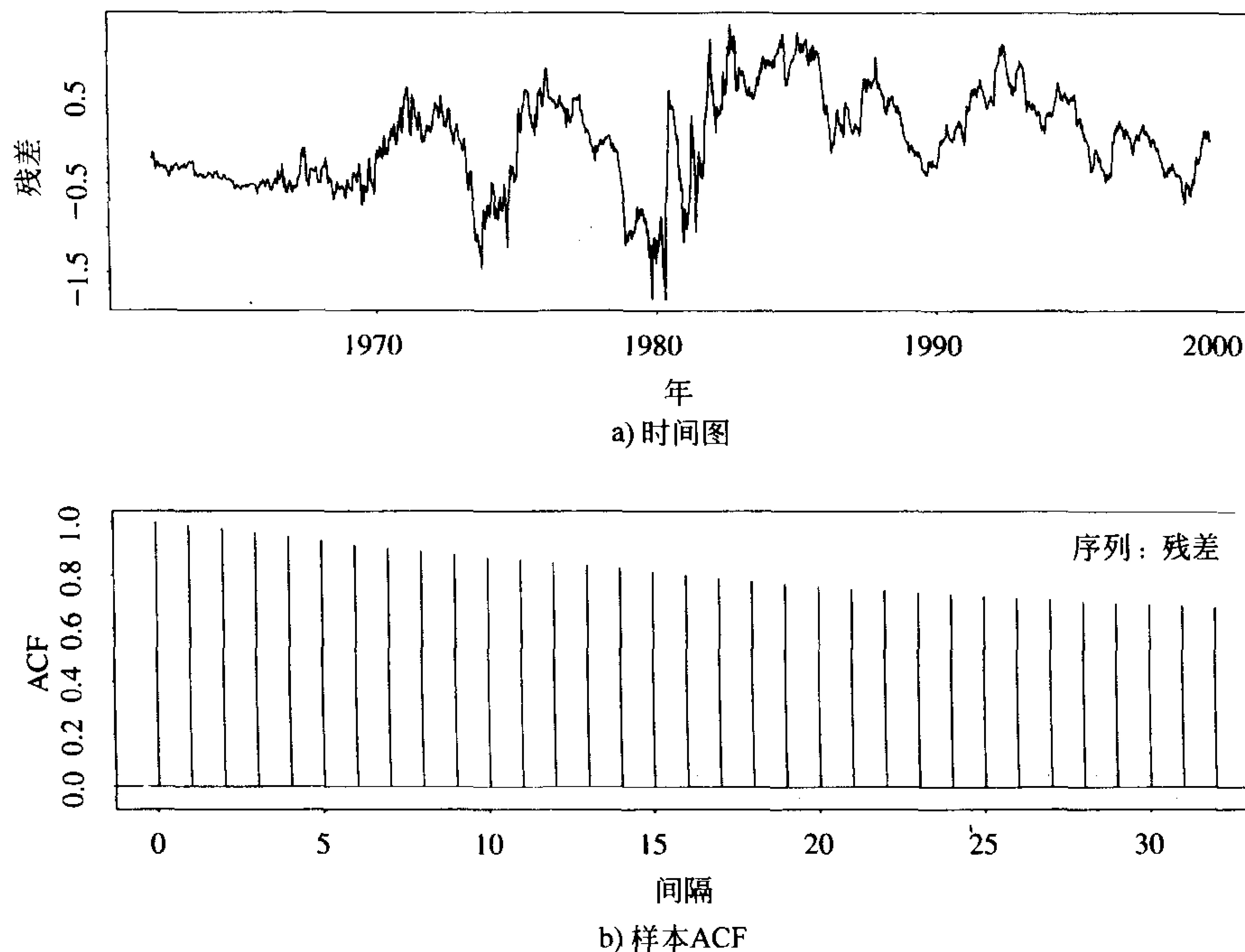


图 2-14 两种美国周利率的线性回归式(2.40)的残差序列

两个利率序列和式(2.40)中的残差的单位根非平稳性使我们想到考虑利率变化序列. 令

1. $c_{1t}=r_{1t}-r_{1,t-1}=(1-B)r_{1t}$, $t\geqslant 2$: 1 年期利率的变化;

2. $c_{3t}=r_{3t}-r_{3,t-1}=(1-B)r_{3t}$, $t\geqslant 2$: 3 年期利率的变化,

并考虑线性回归 $c_{3t}=\alpha+\beta c_{1t}+e_t$. 图 2-15 所示的是两个变化序列的时间图，而图 2-13b 所示的是它们之间的散点图. 这两个变化序列仍然是高度相关的，给它们拟合的线性回归模型为

$$c_{3t}=0.0002+0.7811c_{1t}+e_t,\quad \hat{\sigma}_e=0.0682, \tag{2.41}$$

其 $R^2=84.8\%$，两上系数的标准差分别是 0.0015 和 0.0075. 此模型进一步证实了两种利率间强烈的线性相关性. 图 2-16 所示的是式(2.41)的残差的时间图和样本 ACF，这个样本 ACF 又显示残差中有一些显著的前后相关性，但相关系数的绝对值要小得多. 残差的这种弱前后相关性能用前面几节中讨论的简单时间序列模型来描述，从而得到一个带时间序列误差的线性回归.

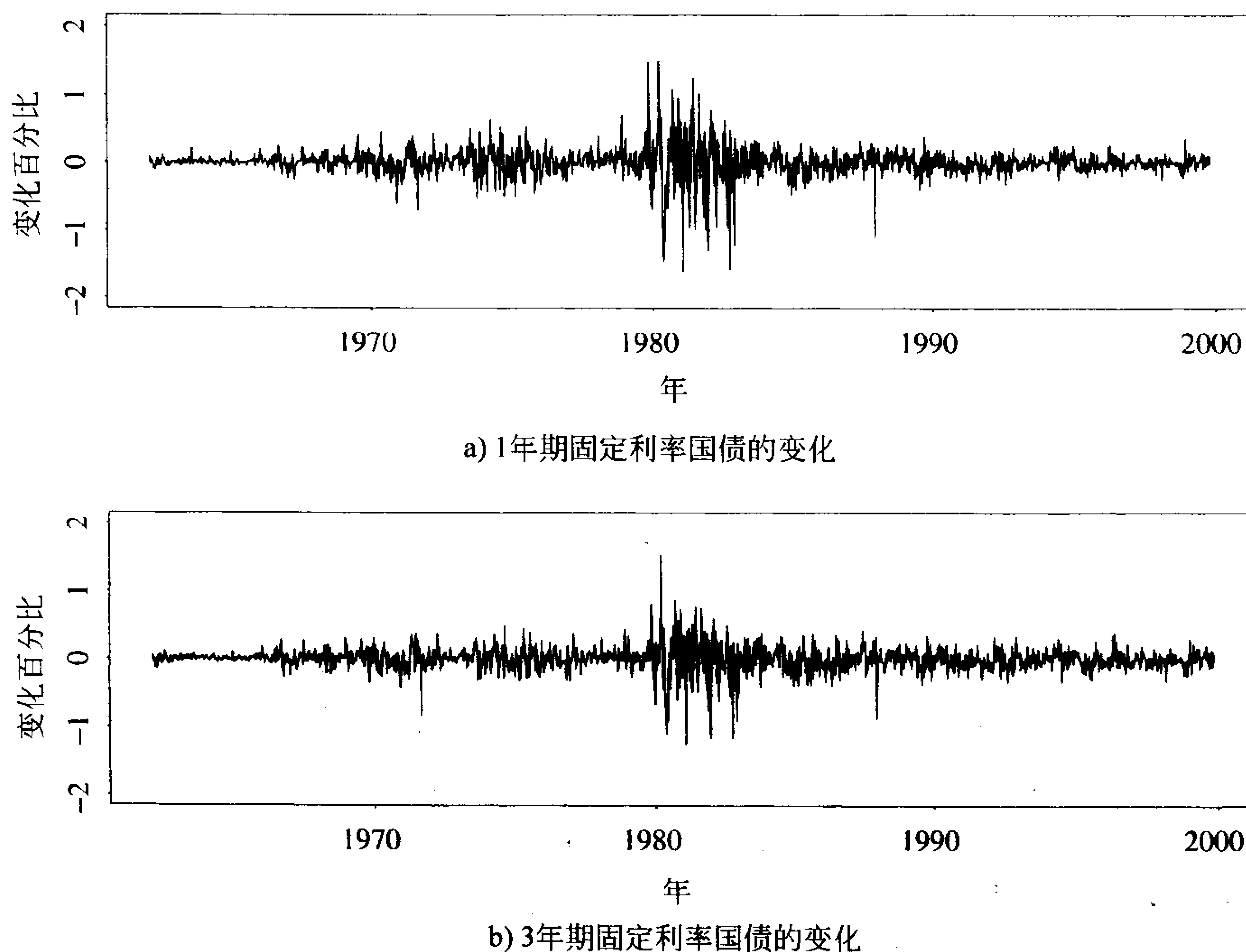


图 2-15 从 1962 年 1 月 12 日到 1999 年 9 月 10 日美国周利率变化序列的时间图

这一节的主要任务是讨论一个简单方法来建立带时间序列误差的线性回归模型. 对残差序列选用一个本章讨论的简单的时间序列模型，然后把整个模型一起估计出来. 这种方法是直接而有效的. 作为说明，考虑式(2.41)中的线性回归. 因为模型的残差是前后相关的，我们给残差识别一个简单的 ARMA 模型. 从图 2-16 的样本 ACF 看出，MA(1)模型对残差序列是适合的，把线性回归模型修改为

$$c_{3t} = \alpha + \beta c_{1t} + e_t, \quad e_t = a_t - \theta_1 a_{t-1}, \tag{2.42}$$

其中$\{a_t\}$是一个白噪声序列. 换句话说，我们简单地用不带常数项的 MA(1)模型来刻画式(2.41)中的误差项的前后相关性. 结果得到的是带时间序列误差的线性回归模型的简单例子. 在实际应用中，可把更复杂的时间序列模型加到线性回归方程上，形成一般的带时间序列误差的回归模型.

67 ~ 70

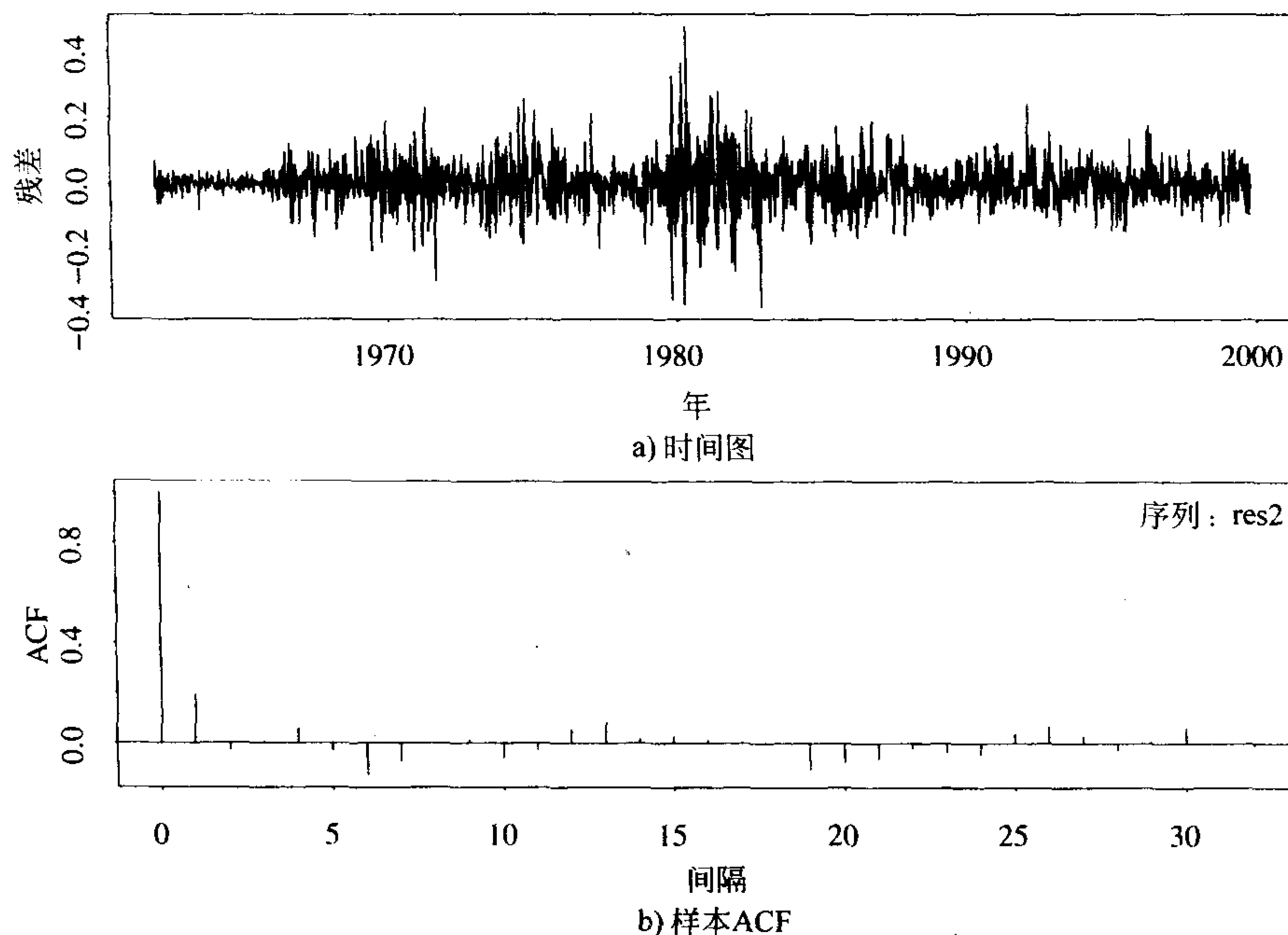


图 2-16 两种美国周利率变化序列的线性回归式(2.41)的残差序列

在现代计算机出现之前，估计带时间序列误差的回归模型不是一件容易的事. 人们提出一些特殊的方法如 Cochrane-Orcutt 估计量，来处理残差中的前后相关性(见 Greene(2000)第 546 页). 但现在这种估计与其他时间序列模型的估计一样容易. 若所用的时间序列模型是平稳的、可逆的，则可用最大似然法把模型一起估计出来. 这就是我们应用 SCA 软件包所采用的方法. 对美国的周利率数据，所拟合的形如式(2.42)的模型为

$$c_{3t} = 0.0002 + 0.7824 c_{1t} + e_t, \quad e_t = a_t + 0.2115 a_{t-1}, \quad \hat{\sigma}_a = 0.0668 \tag{2.43}$$

其 $R^2 = 85.4\%$，参数的标准差分别为 0.0018，0.0077 和 0.0221，此模型不再有显著的间隔为 1 的残差 ACF，虽然在间隔为 4 和 6 时有较小的残差的前后相关系数，但是，如果在残差方程中加上间隔为 4 和 6 的 MA 参数，结果的改进并不大. 具体细节不在这里给出了.

比较式(2.40)、式(2.41)和式(2.43)这三个模型，我们观察到如下几点：首先，式(2.40)的高 R^2 和系数 0.924 会导致人们的错误认识，因为该模型的残差有强烈的前后相关性. 第二，对利率变化序列，式(2.41)和式(2.43)的模型 R^2 和 c_{1t}的系数都很接近，这说明对这个具体例子来说，对变化序列加上一个 MA(1)模型得到不多的改进. 这一点不出乎意料，因为估计出

的 MA 系数不是较大的数值，尽管它是统计显著的. 第三，分析表明在线性回归分析中检验残差的前后相关性是重要的.

因为式(2.43)中的常数项不是显著的，该模型表明前述两个周利率序列有如下关系：

$$r_{3t} = r_{3,t-1} + 0.782(r_{1t} - r_{1,t-1}) + a_t + 0.212a_{t-1}.$$

这两种利率是相互影响并前后相关的.

总结

我们概括一下对带时间序列误差的线性回归模型进行分析的一般方法：

1)拟合一个线性回归模型并检验其残差的前后相关性；

2)如果残差序列是单位根非平稳的，则对响应变量和解释变量都作一阶差分. 然后对两个差分后的序列进行第 1 步. 若这时的残差序列是平稳的，则对残差识别一个 ARMA 模型并相应地修改线性回归模型.

3)用最大似然法进行联合估计，并对拟合模型进行检验，看是否需要进一步改进.

71

检验残差的前后相关性，我们推荐使用 Ljung-Box 统计量，而不是 Durbin-Watson(DW) 统计量，因为后者只考虑间隔为 1 的前后相关性. 有时残差的前后相关性表现在高阶间隔上，尤其是在所涉及的时间序列呈现某种季节性的时候.

注释： *设有残差序列 e_t 的 T 个观察值，Durbin-Watson 统计量是*

$$\mathrm{DW} = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

直接计算表明 $\mathrm{DW} \approx 2(1-\hat{\rho}_1)$，其中 $\hat{\rho}_1$ 是 $\{e_t\}$ 的间隔为 1 的 ACF.

2.10 长记忆模型

我们已经讨论过，平稳时间序列的 ACF 在间隔增加时呈指数速度衰减. 但是，对单位根非平稳时间序列，可以证明：对任意固定的间隔，当样本容量增加时，样本 ACF 收敛于 1.(见 Chan 和 Wei(1988)、Tiao 和 Tsay(1983)). 有些时间序列的 ACF 随间隔的增加是以多项式的速度缓慢衰减到 0 的，这样的过程称为长记忆时间序列. 长记忆序列的一个例子就是如下定义的分数差分序列：

$$(1-B)^d x_t = a_t, \qquad -0.5 < d < 0.5 \tag{2.44}$$

其中 $\{a_t\}$ 是一个白噪声序列. 模型(2.44)的性质已在文献中得到广泛的研究(如 Hosking (1981))，我们把它的一些性质综述如下：

1)若 $d<0.5$，则 x_t 是弱平稳过程并有无穷阶 MA 的表示：

$$x_t = a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i},$$

其中

$$\psi_k = \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}$$

2)若 $d>-0.5$，则 x_t 是可逆的并有无穷阶自回归表示：

$$x_t = \sum_{i=1}^{\infty} \pi_i x_{t-i} + a_t$$

其中

72

$$\pi_k = \frac{-d(1-d)\cdots(k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!}$$

3)对 $-0.5<d<0.5$，x_t 的 ACF 为

$$\rho_k = \frac{d(1+d)\cdots(k-1+d)}{(1-d)(2-d)\cdots(k-d)}, \qquad k=1,2,\cdots.$$

特别地，$\rho_1=d/(1-d)$，且当 $k\to\infty$时

$$\rho_k \approx \frac{(-d)!}{(d-1)!}k^{2d-1}$$

4)对 $-0.5<d<0.5$，x_t 的偏自相关函数(PACF)为 $\phi_{k,k}=d/(k-d)$，$k=1$，2，….

5)对 $-0.5<d<0.5$，x_t 的谱密度函数 $f(\omega)$(它是 x_t 的 ACF 的傅里叶变换)满足

$$f(\omega)\sim\omega^{-2d}, \quad 当\ \omega\to 0, \tag{2.45}$$

其中 $\omega\in[0, 2\pi]$表示频率.

我们对 $d<0.5$ 时，x_t 的 ACF 的性质特别感兴趣. 这条性质表明 $\rho_k\sim ck^{2d-1}$，以多项式速度衰减，而不是以指数速度衰减. 因此，这样的 x_t 过程称为长记忆时间序列. 式(2.45)中的谱密度函数的特殊特征是当 $\omega\to 0$ 时，谱发散到无穷，而平稳 ARMA 过程的谱密度函数对 $\omega\in[0, 2\pi]$是有界的.

前面我们用到了非整数幂的二项展开式：

$$(1-B)^d = \sum_{k=0}^{\infty}(-1)^k\binom{d}{k}B^k, \quad \binom{d}{k} = \frac{d(d-1)\cdots(d-k+1)}{k!}$$

若分数差分序列 $(1-B)^d x_t$ 服从一个 ARMA(p, q)模型，则称 x_t 为一个 ARFIMA(p, d, q)过程，它是 ARIMA 模型的推广，允许 d 为非整数.

在实际中，如果一个时间序列的样本 ACF 在数值上不大，但衰减得很慢，则该序列可能会有长记忆性. 图 2-17 所示 1962 年 7 月 3 日至 1997 年 12 月 31 日的 CRSP 价值加权指数和相等加权指数的日简单收益率的绝对值序列的样本 ACF 值. 可见 ACF 的数值相对较少，但衰减很慢，甚至在间隔为 300 以后还在 5%的水平下是显著的. 对绝对值收益率序列的样本 ACF 的表现，更多的讨论参见 Ding，Granger 和 Engle(1993). 对式(2.44)中的纯分数差分模型，我们可以用最大似然法或带低频对数周期图的回归方法来估计 d. 长记忆模型在金融文献中受到关注，部分的原因是在连续时间模型中关于分形布朗运动的工作.

73

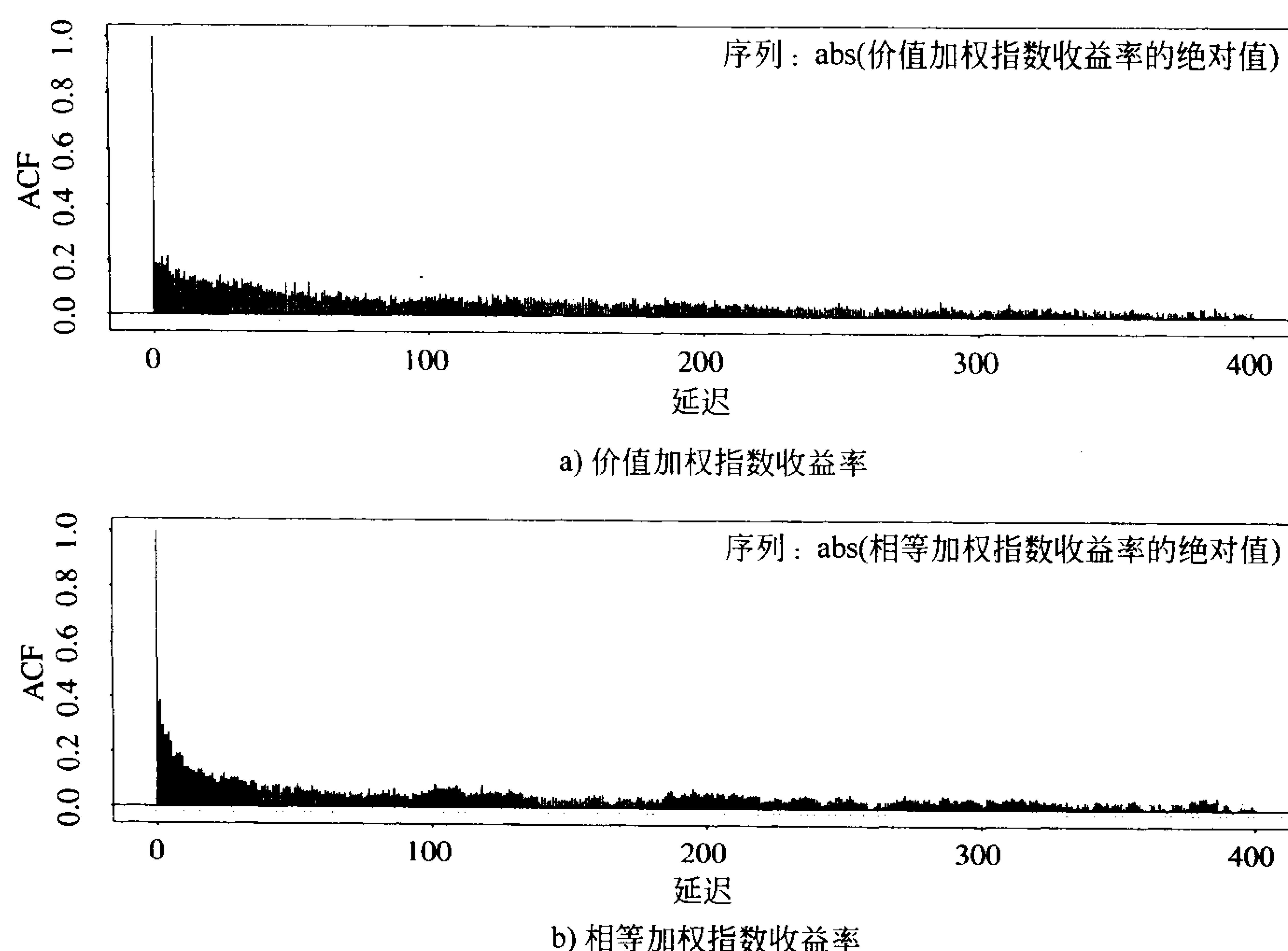


图 2-17　CRSP 价值加权指数和相等加权指数的日简单收益率的绝对值序列的样本 ACF

附录 A　一些 SCA 的命令

A. 2.4 节中使用的命令

数据文件是 m-vw. dat，符号"-"后为注释，这些注释解释的是每个命令的功用.

```
-- load data into SCA and denote the series by vw.
input vw. file 'm-vw.dat'
-- compute 10 lags of PACF.
pacf vw. maxl 10.
-- compute AIC for AR(1) to AR(10).
miden vw. no ccm. arfits 1 to 10.
-- specify an AR(3) model and denote the model by m1.
tsm m1. model (1,2,3)vw=c0+noise.
-- estimate the model and store the residuals in r1.
estim m1. hold resi(r1)
-- compute ACF of the residuals, including Q statistics.
acf r1.
-- refine the model to an AR(5).
tsm m1. model (1,2,3,4,5)vw=c0+noise.
-- estimate the model and store the residuals in r1.
estim m1. hold resi(r1)
-- compute ACF of the residuals.
acf r1. maxl 10.
```

74

```
-- compute p-value of the Q(5) statistic.
p=1.0-cdfc(11.2,5)
-- print p-value.
print p
-- re-estimate the model using the first 858 observations.
estim m1. span 1,858.
-- compute 1-step to 6-step ahead forecasts at origin 858.
ufore m1. orig 858. nofs 6.
-- quit SCA.
stop
```

B. 2.9 节中使用的命令

1 年期限利率在文件“wgs1yr. dat”中，3 年期限利率在文件“wgs3yr. dat”中.

```
-- load the data into SCA, denote the data by rate1 and rate3.
input date, rate1. file 'wgs1yr.dat'
--
input date,rate3. file 'wgs3yr.dat'
-- specify a simple linear regression model.
tsm m1. model rate3=b0+(b1)rate1+noise.
-- estimate the specified model and store residual in r1.
estim m1. hold resi(r1).
-- compute 10 lags of residual acf.
acf r1. maxl 10.
-- difference the two series, denote the new series by c1t and c3t
diff old rate1,rate3. new c1t, c3t. compress.
-- specify a linear regression model for the differenced data
tsm m2. model c3t=h0+(h1)c1t+noise.
-- estimation
estim m2. hold resi(r2).
-- compute residual acf.
acf r2. maxl 10.
-- specify a regression model with time series errors.
tsm m3. model c3t=g0+(g1)c1t+(1)noise.
-- estimate the model using the exact likelihood method.
estim m3. method exact. hold resi(r3).
-- compute residual acf.
acf r3. maxl 10.
-- refine the model to include more MA lags.
tsm m4. model c3t=g0+(g1)c1t+(1,4,6)noise.
-- estimation
estim m4. method exact. hold resi(r4).
-- compute residual acf.
acf r4. maxl 10.
-- exit SCA
stop
```

75

练习题

1. 假定一个债券指数的月简单收益率服从如下 MA(1)模型：

$$R_t = a_t + 0.2a_{t-1}, \quad \sigma_a = 0.025$$

假设 $a_{100}=0.01$. 计算以 $t=100$ 为预测原点的该收益率向前 1 步和向前 2 步的预测. 预测误

差的标准差分别是什么？计算该收益率序列的间隔为 1 和间隔为 2 的自相关系数.

2. 假定一个证券的日对数收益率服从模型

$$r_t = 0.01 + 0.2r_{t-2} + a_t$$

其中$\{a_t\}$是均值为 0、方差为 0.02 的高斯白噪声序列. 收益率序列$\{r_t\}$的均值和方差是什么？计算 r_t 的间隔为 1 和间隔为 2 的自相关系数. 假设 $r_{100}=-0.01$，$r_{99}=0.02$ 计算以 $t=100$ 为预测原点的该收益率向前 1 步和向前 2 步的预测. 预测误差的标准差分别是什么？

3. 文件“bnd.dat”中包含了美国政府的期限为 30 年、20 年、10 年、5 年和 1 年的债券的每月指数的简单收益率数据(以列顺序给出). 数据是由 CRSP 获得的，样本的时间区间是从 1942 年 1 月至 1999 年 12 月. 给期限为 5 年的债券指数的简单收益率建立一个 AR 或 MA 模型. 所拟合的模型是合适的吗？

4. 考虑的样本区间是从 1990 年 1 月至 1999 年 12 月. ALCOA 公司股票的日对数收益率是可预报的吗？你可以检验该假设，通过利用(a)前 5 个间隔的自相关函数的值，和(b)前 10 个间隔的自相关函数的值. 在 5%的显著水平下得出你的结论. 数据可以从 CRSP 得到.

5. 考虑惠普公司股票、价值加权指数、相等加权指数和标准普尔 500 指数从 1980 年 1 月到 1999 年 12 月的日对数收益率，各 5 056 个观察值. 数据是以百分比给出的，可以从 CRSP 或从文件“d-hwp3dx8099.dat”中得到，以列的顺序给出共 4 列. 对每个收益率序列，检验假设 H_0：$\rho_1=\cdots=\rho_{10}=0$ 对备择假设 H_a：$\rho_i\neq 0$，对某 $i\in\{1,\cdots,10\}$，其中 ρ_i 是间隔为 i 的自相关函数. 基于 5%的显著水平得出你的结论. 比较个股和股票指数收益率的结果. [76]

6. 考虑 CRSP 相等加权指数的月对数收益率，从 1962 年 1 月到 1999 年 12 月共 456 个观察值. 你可以直接从 CRSP 或在网上文件“m-ew6299.dat”中得到数据.
   - 给该序列建立一个 AR 模型，并检验所拟合的模型.
   - 给该序列建立一个 MA 模型，并检验所拟合的模型.
   - 比较前两项中所建的 AR 和 MA 模型的向前 1 步和 2 步预测.
   - 比较所拟合的 AR 和 MA 模型.

7. 文件“d-hwp3dx8099.dat”的第 3 列是 CRSP 相等加权指数从 1980 年 1 月到 1999 年 12 月的日对数收益率.
   - 对该序列建立一个 AR 模型，并检验所拟合的模型.
   - 对该序列建立一个 ARMA 模型，并检验所拟合的模型.
   - 利用所建立的 AR 模型，以 1999 年 12 月 27 日为预测原点(即 $h=5\,052$). 计算向前 1 步至向前 7 步预测值. 注意到对这个特殊例子，间隔为 5 的系数是统计显著的. 这可能是由于“周末效应”.

8. 再考虑 CRSP 相等加权指数从 1980 年 1 月到 1999 年 12 月的日对数收益率. 用指示变量代表星期一、星期二、星期三、星期四，用一个回归模型(可能是带时间序列误差的)来研究交易日对指数收益率的效应. 所拟合模型是什么？残差中存在前后相关性吗？

9. 考虑标准普尔 500 指数的现价与期货价的动态关系. 数据文件“sp5may.dat”有三列：期货价的对数，现价的对数，交易成本(cost-of-carry)($\times 100$). 数据是来自芝加哥商品交易所(Chicago Mercantile Exchange)的 1993 年 5 月的标准普尔 500 指数及其 6 月的期货合约，时间间隔为 1 分钟(一天内). 有些作者利用此数据来研究指数期货的套利机会. 这里我们把

注意力放在头两列上．设 f_t 和 s_t 分别是期货价和现价的对数．考虑 $y_t=f_t-f_{t-1}$ 和 $x_t=s_t-s_{t-1}$，设 y_t 为响应变量，建立 $\{y_t\}$ 和 $\{x_t\}$ 间的带时间序列误差的回归模型．

10. 数据文件"qunemrate. dat"包含的是从 1948 年到 1991 年第二季度的、经过季节性调整的美国每季度的失业率．考虑变化序列 $y_t=x_t-x_{t-1}$，其中 x_t 是每季度的失业率．对 y_t 序列建立一个 AR 模型．所拟合的模型显示出商业环的存在吗？ [77]

11. 每季度国民总产值暗含的通货紧缩(gross domestic product implicit price deflator)通常用作为通货膨胀的一个变量．文件"gdpipd. dat"包含了美国从 1947 年第一季度到 2000 年最后一季度的数据．该数据是经过季节性调整的，1996 年时的数值为 100．对该序列建立一个 ARIMA 模型，并检验所拟合的模型的有效性．数据是从圣・路易斯联邦储备银行得到的．

参考文献

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in B.N. Petrov and F. Csaki, ed. *2nd International Symposium on Information Theory*, 267–281. Akademia Kiado: Budapest.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control*, 3rd edition, Prentice Hall: Englewood Cliffs, New Jersey.

Box, G. E. P., and Pierce, D. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–1526.

Brockwell, P. J., and Davis, R. A. (1996), *Introduction to Time Series and Forecasting*, Springer: New York.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd edition, Springer-Verlag: New York.

Chan, N. H., and Wei, C. Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367–401.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimates for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 427–431.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993), "A Long Memory Property of Stock Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, Wiley: New York.

Greene, W. H. (2000), *Econometric Analysis*, 4th edition, Prentice-Hall: Upper Saddle River, New Jersey.

Hosking, J. R. M. (1981), "Fractional Differencing," *Biometrika*, 68, 165–176.

Ljung, G., and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 66, 67–72.

Phillips, P. C. B. (1987), "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

Shumway, R. H., and Stoffer, D. S. (2000), *Time Series Analysis and its Applications*, Springer-Velag: New York.

Tiao, G. C., and Tsay, R. S. (1983), "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models," *Annals of Statistics*, 11, 856–871.

Tsay, R. S., and Tiao, G. C. (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models," *Journal of the American Statistical Association*, 79, 84–96. [78]

第3章　条件异方差模型

本章的目标是研究一些在文献中用来给资产收益率的波动率(volatility)建模的经济计量模型，称为条件异方差(conditional heteroscedastic)模型.

波动率在期权交易中是一个重要因素，它是标的资产收益率的条件方差. 例如，考虑一个欧氏看涨期权的价格. 欧氏看涨期权是一个合同，它给持有人一个权利，可以在将来给定日期以固定的价格购买固定数量的一个具体股票的股份，持有人在对自己不利的情况可以不执行购买. 这个固定的价格叫做敲定价格，通常记为 K. 给定的日期叫做到期日. 这里重要的是距到期日的时间长度，记为 l. 如果持有人能在到期日或者在到期日之前任何时刻执行他的权利，那么这种期权叫做美氏看涨期权. 著名的 Black-Scholes 期权定价公式表明一个欧氏看涨期权的价格是

$$c_t = P_t\Phi(x) - Kr^{-l}\Phi(x - \sigma_t\sqrt{l}), \quad x = \frac{\ln(P_t/Kr^{-l})}{\sigma_t\sqrt{l}} + \frac{1}{2}\sigma_t\sqrt{l} \tag{3.1}$$

其中 P_t 是标的股票的现价，r 是无风险利率，σ_t 是这个股票的对数收益率的条件标准差，$\Phi(x)$是标准正态随机变量的累积分布函数在 x 点的值. 此公式的推导将在第 6 章给出. 对这个公式有好几个精彩解释，但在这里只要看到标的资产对数收益率的条件方差在其中起主要作用这一点就够了. 这个波动率是随时间变化的.

在风险管理中波动率也是很重要的. 如第 7 章的讨论，波动率的建模给计算一个金融头寸的风险值(value at risk)提供一个简单方法. 最终一个时间序列波动率的建模能改进参数估计的有效性和区间估计的精确度.

本章讨论的一元波动率模型包括 Engle 在 1982 年提出的自回归条件异方差(ARCH)模型，Bollerslev 在 1986 年提出的推广的自回归条件异方差模型(GARCH)，Nelson 在 1991 年提出的指数 GARCH 模型(EGARCH)，Tsay 在 1987 年提出的条件异方差自回归滑动平均(CHARMA)模型，Nicholls 和 Quinn 在 1982 年提出的随机系数自回归(RCA)模型，Melino 和 Turnbull 在 1990 年，Harvey，Ruiz 和 Shephard 在 1994 年，Jacquier，Polson 和 Rossi 在 1994 年分别提出的随机波动率(SV)模型. 我们也讨论各个波动率模型的优点和缺点，给出这些模型的一些应用. 多元波动率模型，包括那些具有时变相关系数的模型，将在第 9 章中讨论.

3.1　波动率的特征

股票波动率的一个特殊性是它不能被直接观测. 例如，考虑 IBM 股票的日对数收益率. 因为一个交易日只有一个观测值，所以日波动率是不能从收益率中观测出来的. 如果有一天内的股票数据(如 5 分钟的收益率)，那么我们可估计日波动率，但这样估计的精确度应予小心研

究．另外，股票波动率包括一天内的波动率和不同交易日之间的变化两个方面．波动率的不可观测性给评价条件异方差模型的预测表现带来了困难．我们在后面将讨论这一问题．

在期权市场上，如果我们接受这样一个思想，期权的价格是由如 Black-Scholes 公式这样的计量经济模型决定的，那么我们可以利用期权的价格得到“隐含”波动率(implied volatility)．这种方法往往遭到批评，因为使用的具体模型，要基于一些实际可能不成立的假定．例如，由一个欧氏看涨期权的价格，我们能利用等式(3.1)的 Black-Scholes 公式去推导出条件标准差 σ_t，所得到的值 σ_t^2 称为标的股票的隐含波动率．然而，这个隐含波动率是在对收益率序列的对数正态的假定下得到的，它可能与实际的波动率很不一样．经验告诉我们，资产收益率的隐含波动率一般要比采用 GARCH 类型的波动率模型得到的值大．

虽然波动率不能被直接观测，但它的一些特征往往是在资产收益率序列中能看到的．首先，存在波动率聚类(volatility cluster)(也就是，波动率可能在一些时间段上高，而在另一些时间段上低)．第二，波动率以连续方式随时间变化，即波动率跳跃是很少见的．第三，波动率不发散到无穷，即波动率是在固定的范围内变化．从统计学角度说，这意味着波动率往往是平稳的．第四，波动率对价格的大幅上升和价格的大幅下降的反应不同．这些性质对波动率模型的发展起着重要作用．为了克服现存模型在刻画上述这些特征上的弱点，提出了一些新的模型．例如，EGARCH 模型就是为了刻画波动率对大的“正”和“负”资产收益率的不对称性而提出来的．

80

3.2 模型的结构

用 r_t 表示资产在 t 时刻的对数收益率．波动率研究的基本思想是，序列 $\{r_t\}$ 是前后不相关的或低阶前后相关的，但不是独立的．图 3-1 是 Intel 公司股票从 1973 年 1 月到 1997 年 12 月的每月对数收益率的自相关函数(ACF)和偏自相关函数(PACF)．左上角的是收益率的样本 ACF，显示除了在间隔为 7 时有较小相关性之外没有显著的前后相关性．右上角是对数收益率的绝对值 $|r_t|$ 的样本 ACF，而左下角是收益率平方 r_t^2 的样本 ACF，这两幅图清楚地表明月收益率序列不是独立的．结合上述三幅图，收益率序列似是前后不相关的，但不是独立的．波动率模型就试图刻画收益率序列这种不独立性．

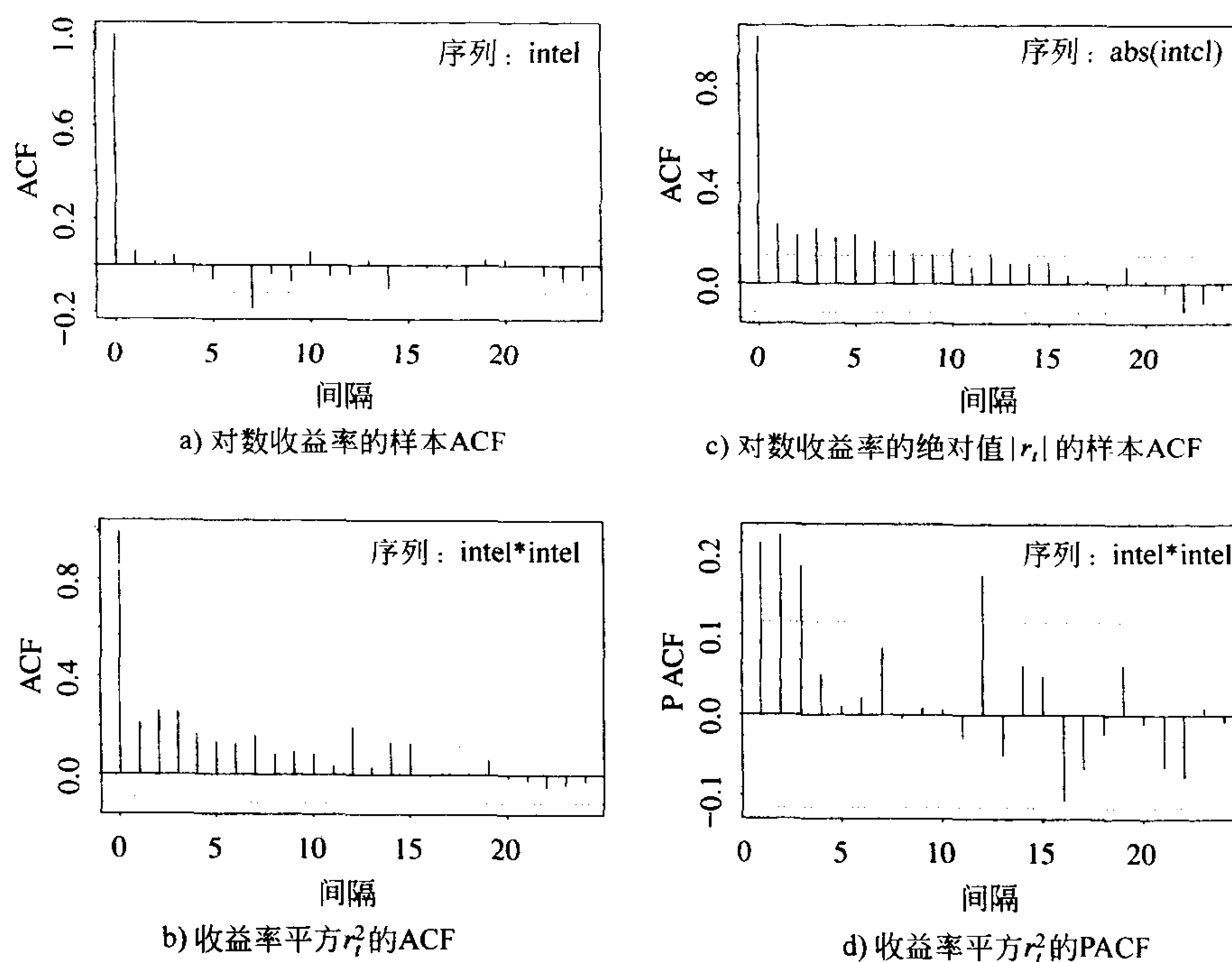
为了把波动率模型放在一个合理的框架中，考虑给定 $t-1$ 时刻已知的信息集 F_{t-1} 时，r_t 的条件均值和条件方差，即

$$\mu_t = E(r_t \mid F_{t-1}), \quad \sigma_t^2 = \mathrm{Var}(r_t \mid F_{t-1}) = E[(r_t - \mu_t)^2 \mid F_{t-1}] \tag{3.2}$$

是有益的．信息集 F_{t-1} 典型的取法是由过去收益率的全体线性函数组成．第 2 章中的实例和图 3-1 表明：股票收益率序列 r_t 具有很弱的前后相关性．因此，式(3.2)中 μ_t 的等式应是较简单的，我们假定 r_t 服从一个简单的时间序列模型，如平稳 ARMA(p, q)模型．换句话说，对于 r_t 我们接受模型

$$r_t = \mu_t + a_t, \quad \mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}, \tag{3.3}$$

其中 p 和 q 是非负整数．



a) 对数收益率的样本ACF

c) 对数收益率的绝对值 $|r_t|$ 的样本ACF

b) 收益率平方 r_t^2 的ACF

d) 收益率平方 r_t^2 的PACF

图 3-1 Intel 公司股票从 1973 年 1 月到 1997 年 12 月的每月对数收益率的样本自相关函数(ACF)和样本偏自相关函数(PACF)

模型(3.3)说明第 2 章中线性时间序列模型在金融中的可能应用. ARMA 模型的阶 (p, q) 可能取决于收益率序列的频率. 例如，股票市场指数的日收益率往往有较小的前后相关性，但指数月收益率可能就没有任何显著的前后相关性. 我们可以对条件均值方程加入一些解释变量，利用带时间序列误差项的线性回归模型来刻画 μ_t 的性质. 例如，对星期一这天设计一个“哑变量”来研究股票日收益率的周末效应.

结合式(3.2)和式(3.3)，我们有

$$\sigma_t^2 = \mathrm{Var}(r_t \mid F_{t-1}) = \mathrm{Var}(a_t \mid F_{t-1}). \tag{3.4}$$

本章的条件异方差模型是用来描述 σ_t^2 的演变的. σ_t^2 随时间变化的方式可以用不同的波动率模型来表示.

条件异方差模型可分为两类：第一类是用确定的函数来刻画 σ_t^2 的演变，第二类是用随机方程来描述 σ_t^2. GARCH 模型属于第一类，而随机波动率模型属于第二类.

在介绍波动率模型时，为简单起见，我们假定模型的条件均值是给定的. 然而，在实证分析中我们可同时估计条件均值和条件方差的方程. 全书中把 a_t 称为资产收益率在 t 时刻的扰动或均值修正后的收益率，σ_t 为 σ_t^2 的正平方根. 式(3.3)中 μ_t 的模型叫做 r_t 均值方程，σ_t^2 的模型叫做 r_t 的波动率方程. 因此，对条件异方差性建模就是要用一个时间序列模型的动态方程，

来刻画抖动的条件方差的时间演变规律.

3.3 ARCH 模型

81 ~ 82

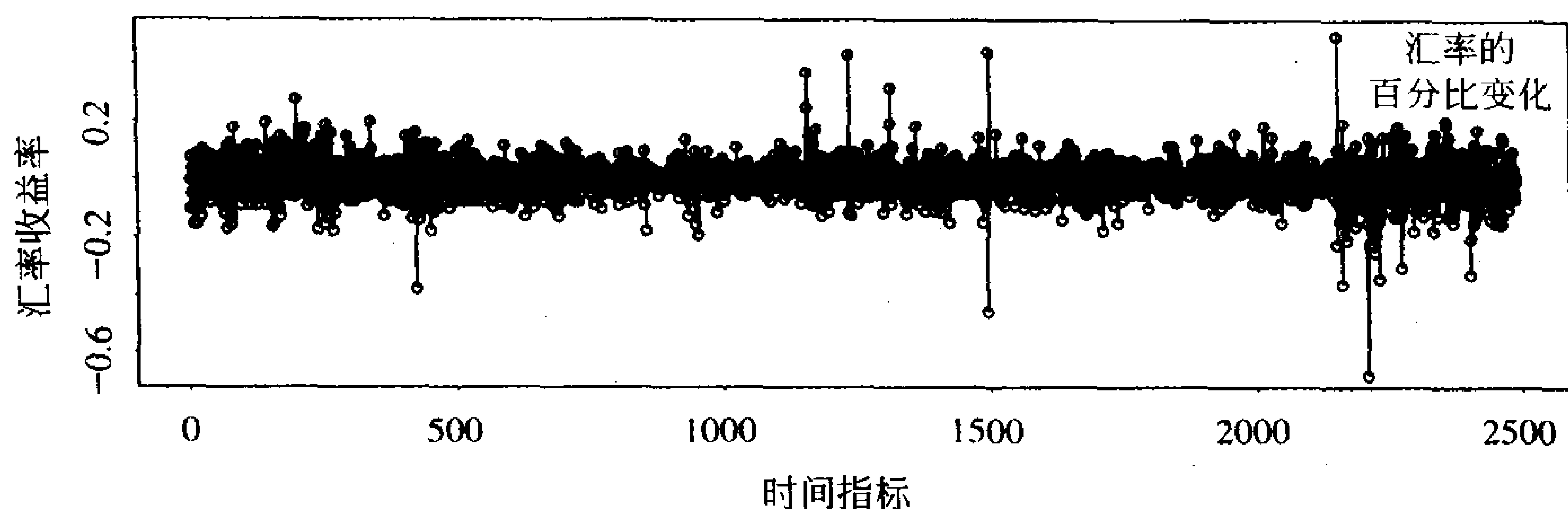
给波动率建模提供一个系统框架的第一个模型是 Engle 在 1982 年提出的 ARCH 模型. ARCH 模型的基本思想是:(a)均值修正的资产收益率 a_t 是前后不相关的,但不是独立的;(b)a_t 的不独立性可以用一个它的延迟值的简单二次函数来描述. 具体地说,一个 ARCH(m)模型假定

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_m a_{t-m}^2 \tag{3.5}$$

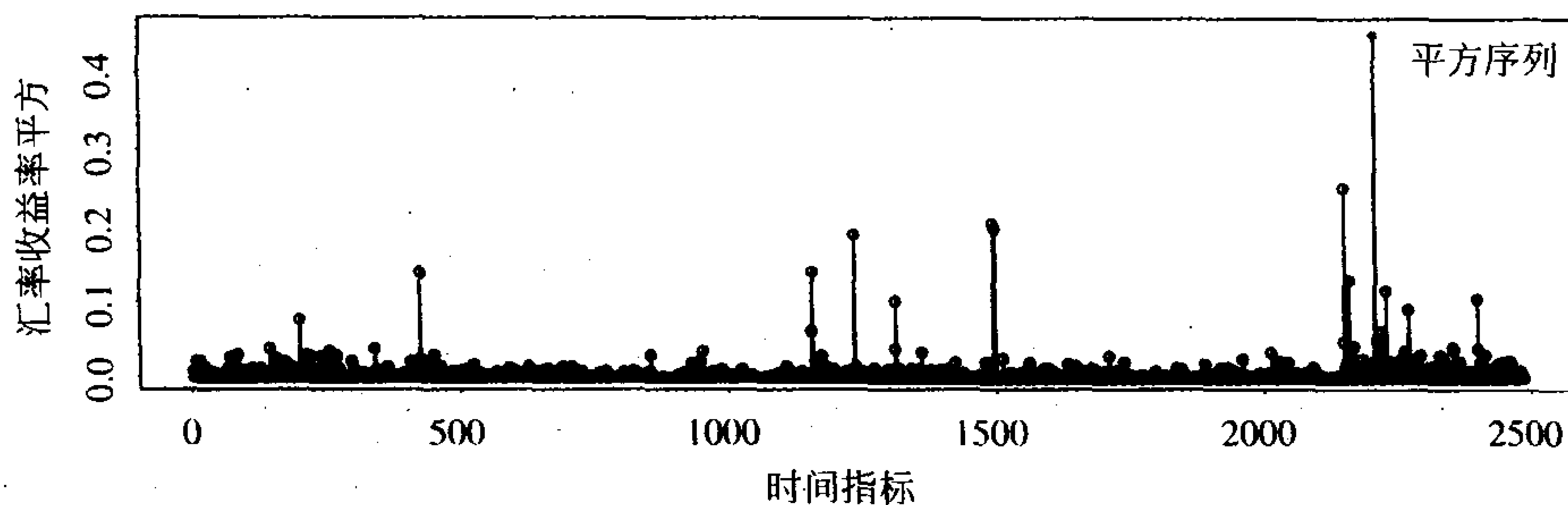
其中$\{\varepsilon_t\}$是独立同分布(iid)的随机变量序列,均值为 0,方差为 1,$\alpha_0>0$,对$i>0$有 $\alpha_i \geqslant 0$. 系数 α_i 必须满足一些正则性条件保证 a_t 的无条件方差是有限的. 实际中,通常假定 ε_t 服从标准正态分布或标准化的学生-t 分布.

从模型的结构上看,大的过去抖动的平方$\{a_{t-i}^2\}_{i=1}^m$导出均值修正的收益率 a_t 的大的条件方差σ_t^2. 从而,a_t 有(取绝对值)较大的值的倾向. 这意味着,在 ARCH 的框架下,大的抖动会接着另一个大的抖动. 这里用倾向这个词,是因为大方差不一定意味着大的变化量,它只表明是大的变化发生的概率比小方差情形要大. 这种现象与对资产收益率所观察到的波动率聚集相似.

ARCH 效应也出现在其他金融时间序列中. 图 3-2a 显示 1989 年 6 月 5 日至 1989 年 6 月 19 日 10 分钟间隔观察的德国马克对美元的汇率的百分比变化率(2488 个观测值),图 3-2b 是



a) 1989年6月5日至1989年6月19日10分钟间隔的德国马克对美元汇率的收益率

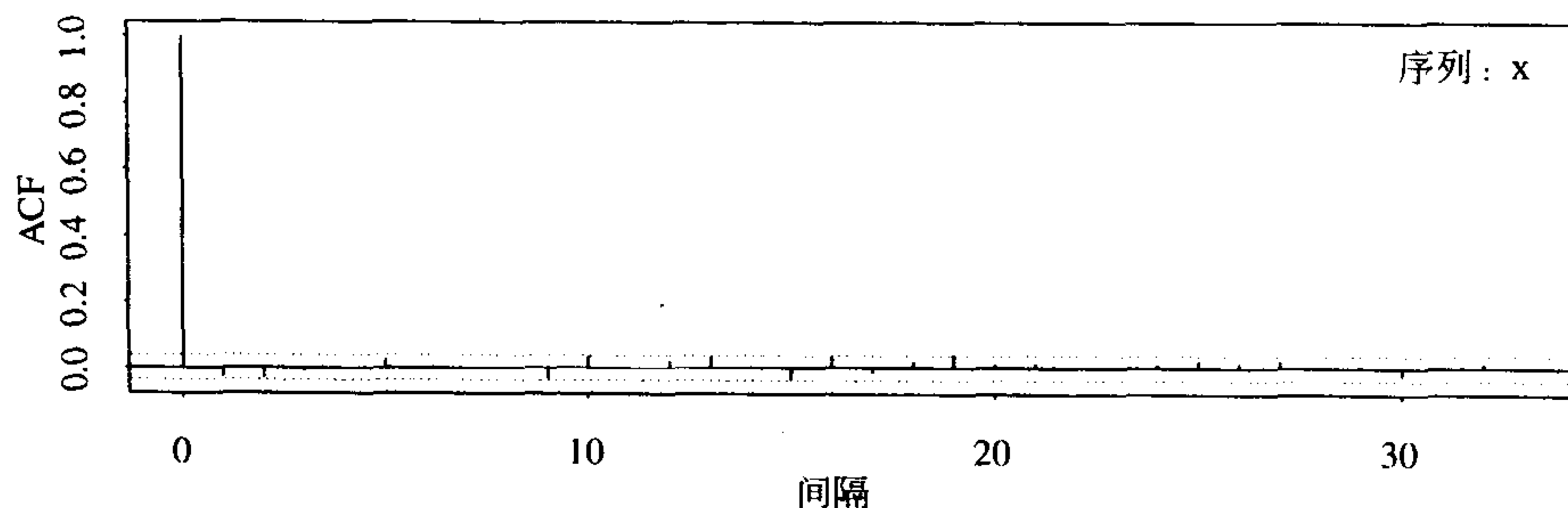


b) 收益率的平方

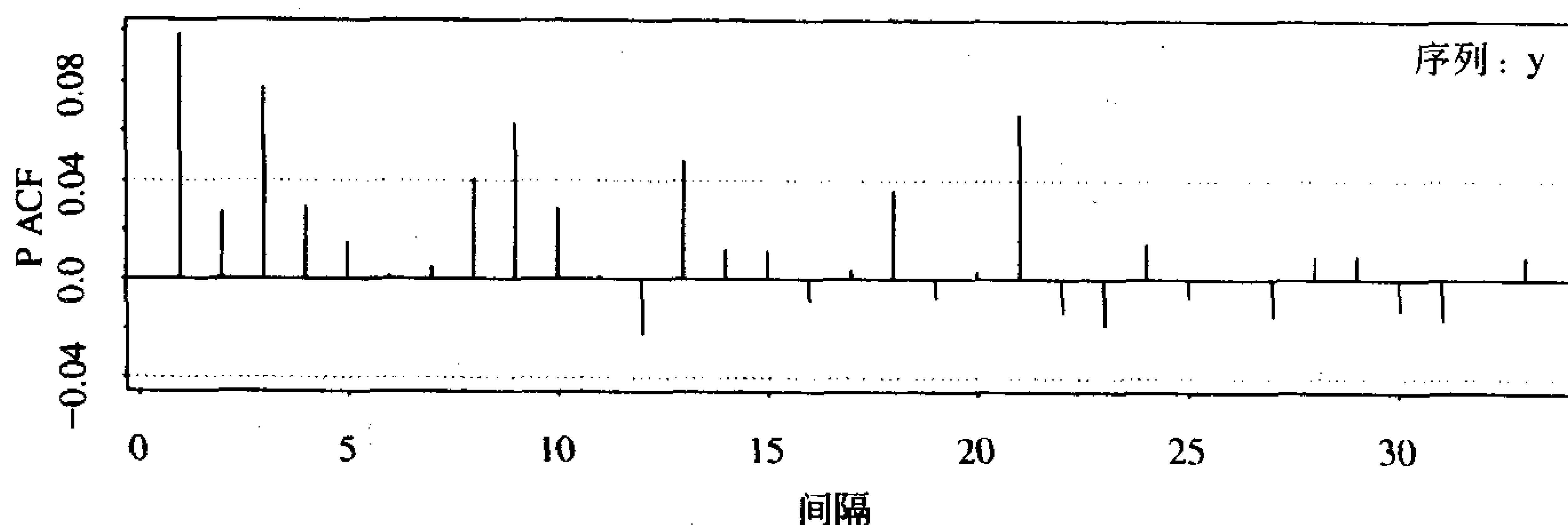
图 3-2

这个百分比变化率的平方所构成的序列. 大的百分比变化率时有发生，但存在某种稳定的周期. 图 3-3a 显示的是这个百分比变化量序列的样本自相关函数，它清楚地表明该序列没有前后相关性. 图 3-3b 是百分比变化率的平方所构成的序列的样本偏自相关函数，可见这个函数有一些大的尖峰，这样的尖峰表明百分比变化率序列不是独立的，有所谓的 ARCH 效应.

注释： 有些作者把式(3.5)中的条件方差表示为 h_t，这时抖动变成 $a_t=\sqrt{h_t}\epsilon_t$.



a) 德国马克对美元的汇率收益率序列的样本自相关函数



b) 平方收益率的样本偏自相关函数

图 3-3

3.3.1 ARCH 模型的性质

为了理解 ARCH 模型的含义，我们来仔细研究一下 ARCH(1)模型：

$$a_t=\sigma_t\epsilon_t,\quad \sigma_t^2=\alpha_0+\alpha_1 a_{t-1}^2$$

其中 $\alpha_0>0$，$\alpha_1\geqslant 0$. 首先，a_t 的无条件均值仍是 0，因为

$$E(a_t)=E[E(a_t\mid F_{t-1})]=E[\sigma_t E(\epsilon_t)]=0$$

其次，a_t 的无条件方差是

$$\mathrm{Var}(a_t)=E(a_t^2)=E[E(a_t^2\mid F_{t-1})]=E(\alpha_0+\alpha_1 a_{t-1}^2)=\alpha_0+\alpha_1 E(a_{t-1}^2).$$

因为 a_t 是平稳过程，且 $E(a_t)=0$，所以 $\mathrm{Var}(a_t)=\mathrm{Var}(a_{t-1})=E(a_{t-1}^2)$. 从而，我们有 $\mathrm{Var}(a_t)=\alpha_0+\alpha_1\mathrm{Var}(a_t)$，$\mathrm{Var}(a_t)=\alpha_0/(1-\alpha_1)$. 因为 a_t 的方差必须为正，我们需要条件 $0\leqslant\alpha_1<1$. 第三，在一些应用中，我们需要 a_t 的更高阶的矩存在，从而，α_1 还需满足另外的

约束条件. 例如，为研究 a_t 的尾部性质，我们要求 a_t 的四阶矩是有限的. 在方程(3.5)中 ε_t 是正态分布的假定下，我们有

$$E(a_t^4 \mid F_{t-1}) = 3[E(a_t^2 \mid F_{t-1})]^2 = 3(\alpha_0 + \alpha_1 a_{t-1}^2)^2$$

因此

$$E(a_t^4) = E[E(a_t^4 \mid F_{t-1})] = 3E(\alpha_0 + \alpha_1 a_{t-1}^2)^2 = 3E[\alpha_0^2 + 2\alpha_0\alpha_1 a_{t-1}^2 + \alpha_1^2 a_{t-1}^4]$$

若 a_t 是四阶平稳的，且 $m_4 = E(a_t^4)$，则我们有

$$\begin{aligned} m_4 &= 3[\alpha_0^2 + 2\alpha_0\alpha_1 \mathrm{Var}(a_t) + \alpha_1^2 m_4] \\ &= 3\alpha_0^2\left(1 + 2\frac{\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2 m_4 \end{aligned}$$

从而，

$$m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

这个结果有两个重要含义：(a)因为 a_t 的四阶矩是正的，所以 α_1 必须满足 $1-3\alpha_1^2>0$，即 $0\leqslant\alpha_1^2<1/3$；(b)a_t 的无条件峰度是

$$\frac{E(a_t^4)}{[\mathrm{Var}(a_t)]^2} = 3\frac{\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3$$

这样，a_t 的超出峰度是正的，并且 a_t 的分布尾部比正态分布的尾部厚. 换句话说，服从条件高斯的 ARCH(1)模型的扰动 a_t 比高斯白噪声序列更容易产生“异常值”(outliers)，这是与经验结果相一致的，经验结果也表明资产收益率中出现“异常值”的时候要比独立同分布的正态随机变量序列出现“异常值”的时候多.

83～85

这些性质对一般的 ARCH 模型仍成立，但对高阶 ARCH 模型公式会变得更复杂一些. 式(3.5)中条件 $\alpha_i\geqslant0$ 能放松，这是一个保证条件方差 σ_t^2 对所有的 t 取正值的条件. 事实上，使条件方差取正值的一个自然方式是把 ARCH(m)模型改写成

$$a_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = \alpha_0 + A'_{m,t-1}\Omega A_{m,t-1} \tag{3.6}$$

其中 $A_{m,t-1}=(a_{t-1}, \cdots, a_{t-m})'$，$\Omega$ 是一个 $m\times m$ 阶的非负定矩阵. 式(3.5)定义的 ARCH(m)模型要求 Ω 是对角阵. 这样，Engle 的模型是用一个非常节省参数的方式来逼近一个二次函数. 达到式(3.6)的一个简单方法是给 a_t 一个随机系数模型(参见后面讨论的 CHARMA 和 RCA 模型).

3.3.2 ARCH 模型的缺点

ARCH 模型有不少优点，包括前一节中讨论的性质. 这个模型也有一些缺点：

1)ARCH 模型假定正的扰动和负的扰动对波动率有相同的影响，因为波动率依赖于以前扰动的平方. 实际中，众所周知，金融资产的价格对正的和负的扰动的反应是不同的.

2)ARCH 模型对参数的限制是相当严格的. 比如，若序列有有限的四阶矩，则 ARCH(1)模型中的 α_1^2 必须在$\left[0, \frac{1}{3}\right]$区间中. 对高阶的 ARCH 模型，这种约束会变得更复杂.

3)对于弄清一个金融时间序列的变化的来源，ARCH 模型不能提供任何新见解. 它只是

提供一个机械的方式来描述条件方差的状态，而对由什么引起这种变化没有给出任何启示.

4)ARCH 模型会过高估计波动率，因为它对收益率序列大的孤立的抖动反应缓慢.

3.3.3 ARCH 模型的建立

建立一个 ARCH 模型的简单方法包括三个步骤：(1)对收益率序列建立一个经济计量模型(如 ARMA 模型)，以分离出数据中的任何线性相关成分，并用该模型的残差序列检验 ARCH 效应；(2)具体确定 ARCH 模型的阶，并估计参数；(3)仔细检验所拟合的 ARCH 模型，对它进行必要的改进. 更多细节将在后面讨论.

对均值效应建立模型和检验

对所观察到的时间序列建立一个 ARMA 模型，分离出数据中的先后相关性. 对大多数资产收益率序列来说，这一步就是把样本均值从数据中分离出来，如果样本均值显著地不同于零. 对某些日收益率序列，可能会需要一个简单的 AR 模型. 设 $a_t=r_t-\mu_t$ 是该 ARMA 模型的残差，用平方序列 a_t^2 来检验条件异方差性. 有两种检验方法可供使用. 第一种检验方法是验证 a_t^2 的通常 Ljung-Box 统计量，(见 McLeod 和 Li(1983)). 第二种方法是 Engle 在 1982 年提出的拉格朗日乘子法，这种检验法等价于通常用来检验关于线性回归 [86]

$$a_t^2=\alpha_0+\alpha_1 a_{t-1}^2+\cdots+\alpha_m a_{t-m}^2+e_t,\qquad t=m+1,\cdots,T,$$

的假设 $\alpha_i=0(i=1,\cdots,m)$的 F 检验法，其中 e_t 表示误差项，m 是一个事先给定的正整数，T 是样本容量. 令 $SSR_0=\sum_{t=m+1}^{T}(a_t^2-\bar{\omega})^2$，其中$\bar{\omega}$表示 a_t^2 的样本均值，$SSR_1=\sum_{t=m+1}^{T}\hat{e}_t^2$，$\hat{e}_t$ 是上述线性回归的最小二乘残差. 我们有

$$F=\frac{(SSR_0-SSR_1)/m}{SSR_1/(T-2m-1)}.$$

在原假定 $\alpha_i=0$，$i=1,\cdots,m$ 成立时，这个 F 统计量是渐近的自由度为 m 的 χ^2 分布的.

阶的确定

如果检验统计量 F 是显著的，则认为 a_t 有条件异方差性，我们用 a_t^2 的偏自相关函数来确定 ARCH 的阶. 从式(3.5)定义的模型出发，我们有

$$\sigma_t^2=\alpha_0+\alpha_1 a_{t-1}^2+\cdots+\alpha_m a_{t-m}^2$$

对给定的样本，a_t^2 是 σ_t^2 的无偏估计. 因此，我们期望 a_t^2 以 m 阶自回归模型的类似方式与 $a_{t-1}^2,\cdots,a_{t-m}^2$线性相关. 注意到单个的 a_t^2 往往不是σ_t^2 的有效估计，但它可以作为一个近似，在具体确定阶 m 时将会提供充分的信息.

从另一角度，定义 $\eta_t=a_t^2-\sigma_t^2$. 那么可以证明$\{\eta_t\}$是均值为零的不相关序列. 于是 ARCH 模型变成

$$a_t^2=\alpha_0+\alpha_1 a_{t-1}^2+\cdots+\alpha_m a_{t-m}^2+\eta_t,$$

这是 a_t^2 的 AR(m)形式，但$\{\eta_t\}$不是独立同分布序列. 由第 2 章的内容，a_t^2 的 PACF 是确定阶 m 的有用工具. 因为$\{\eta_t\}$不是同分布的，上述模型的最小二乘估计是相合的，但不是有效的. 当样本容量较小时，a_t^2 的 PACF 可能不是有效的. [87]

估计

在 ARCH 模型的估计中通常使用两个似然函数. 在正态性的假定下，ARCH(m)模型的似然函数为

$$f(a_1,\cdots,a_T \mid \boldsymbol{\alpha}) = f(a_T \mid F_{T-1})f(a_{T-1} \mid F_{T-2})\cdots f(a_{m+1} \mid F_m)f(a_1,\cdots,a_m \mid \boldsymbol{\alpha})$$
$$= \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left[-\frac{a_t^2}{2\sigma_t^2}\right]\times f(a_1,\cdots,a_m \mid \boldsymbol{\alpha}),$$

其中$\boldsymbol{\alpha}=(\alpha_0, \alpha_1, \cdots, \alpha_m)'$，$f(a_1, \cdots, a_m \mid \boldsymbol{\alpha})$是$a_1, \cdots, a_m$ 的联合概率密度函数. 因为$f(a_1, \cdots, a_m \mid \boldsymbol{\alpha})$的精确形式是复杂的，通常把它从上述似然函数中去掉，特别是当样本容量很大时. 这就导出条件似然函数

$$f(a_{m+1},\cdots,a_T \mid \boldsymbol{\alpha}, a_1,\cdots,a_m) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left[-\frac{a_t^2}{2\sigma_t^2}\right]$$

其中 σ_t^2 可以递推地计算. 我们把由上式最大化而得到的估计称为正态假设下条件最大似然估计(MLE).

最大化条件似然函数等价于最大化它的对数，而后者比前者容易处理一些. 条件对数似然函数是

$$l(a_{m+1},\cdots,a_T \mid \boldsymbol{\alpha}, a_1,\cdots,a_m) = \sum_{t=m+1}^{T}\left[-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln(\sigma_t^2)-\frac{1}{2}\frac{a_t^2}{\sigma_t^2}\right]$$

因为第一项 $\ln(2\pi)$不包含任何参数，对数似然函数变成

$$l(a_{m+1},\cdots,a_T \mid \boldsymbol{\alpha}, a_1,\cdots,a_m) = -\sum_{t=m+1}^{T}\left[\frac{1}{2}\ln(\sigma_t^2)+\frac{1}{2}\frac{a_t^2}{\sigma_t^2}\right]$$

其中 $\sigma_t^2=\alpha_0+\alpha_1 a_{t-1}^2+\cdots+\alpha_m a_{t-m}^2$可递推地计算.

在有些应用中，假设 ϵ_t 服从像标准化的学生-t 分布这样的厚尾分布更合适一些. 设随机变量 x_v 服从自由度为 v 的学生-t 分布，则对 $v>2$ 有 $\mathrm{Var}(x_v)=v/(v-2)$，我们有 $\epsilon_t = x_v/\sqrt{v/(v-2)}$. ϵ_t 的概率密度函数为

88

$$f(\epsilon_t \mid v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}}\left(1+\frac{\epsilon_t^2}{v-2}\right)^{-(v+1)/2}, \quad v>2 \tag{3.7}$$

其中 $\Gamma(x)$ 是通常的伽马(gamma) 函数$\left[\text{即}, \Gamma(x)=\int_0^\infty y^{x-1}\mathrm{e}^{-y}\mathrm{d}y\right]$. 利用 $a_t=\sigma_t\epsilon_t$，我们得到 a_t 的条件似然函数

$$f(a_{m+1},\cdots,a_T \mid \boldsymbol{\alpha}, A_m) = \prod_{t=m+1}^{T} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{(v-2)\pi}}\frac{1}{\sigma_t}\left[1+\frac{a_t^2}{(v-2)\sigma_t^2}\right]^{-(v+1)/2},$$

其中 $v>2$，$A_m=(a_1, \cdots, a_m)$. 我们把由上式最大化得到的估计称为 t 分布假设下的条件 MLE. t 分布的自由度可以先验地给出，也可以和其他参数一起估计出来，如果是事先给定的，往往是 3 到 6 之间的一个值.

如果学生-t 分布的自由度是事先给定的，那么条件对数似然函数为

$$l(a_{m+1},\cdots,a_T \mid \boldsymbol{\alpha}, A_m) = -\sum_{t=m+1}^{T}\left[\frac{v+1}{2}\ln\left(1+\frac{a_t^2}{(v-2)\sigma_t^2}\right)+\frac{1}{2}\ln(\sigma_t^2)\right]. \tag{3.8}$$

如果想把 v 和其他参数一起估计出来，那么包含自由度的对数似然函数为

$$l(a_{m+1},\cdots,a_T \mid \boldsymbol{\alpha},v,A_m) = (T-m)[\ln(\Gamma((v+1)/2)) - \ln(\Gamma(v/2)) - 0.5\ln((v-2)\pi)] + l(a_{m+1},\cdots,a_T \mid \boldsymbol{\alpha},A_m)$$

其中第二项由式(3.8)给出.

模型的验证

对一个 ARCH 模型，标准化的抖动

$$\tilde{a}_t = \frac{a_t}{\sigma_t}$$

是服从标准正态或标准化的学生-t 分布的 iid 随机变量. 因此，我们可通过检查序列$\{\tilde{a}_t\}$来验证所拟合的 ARCH 模型是否合适. 特别是，$\tilde{a}_t$ 的 Ljung-Box 统计量可用来检验均值方程，$\tilde{a}_t^2$ 的 Ljung-Box 统计量可用来检验波动率方程有效性. $\{\tilde{a}_t\}$的偏度、峰度、分位点对分位点图(QQ 图)可用来检验分布假设的有效性.

预测

式(3.5)定义的 ARCH 模型的预测可像 AR 模型一样递推地得到. 考虑一个 ARCH(m)模型. 从预测原点 h 出发，向前 1 步 σ_{h+1}^2 的预测为

89

$$\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + \cdots + \alpha_m a_{h+1-m}^2$$

向前 2 步预测为

$$\sigma_h^2(2) = \alpha_0 + \alpha_1 \sigma_h^2(1) + \alpha_2 a_h^2 + \cdots + \alpha_m a_{h+2-m}^2$$

向前 l 步预测为

$$\sigma_h^2(l) = \alpha_0 + \sum_{i=1}^{m} \alpha_i \sigma_h^2(l-i) \tag{3.9}$$

其中 $\sigma_h^2(l-i) = a_{h+l-i}^2$，对 $l-i \leqslant 0$.

3.3.4 例子

这一小节，我们通过两个例子来说明 ARCH 的建模.

例 3.1 我们首先应用建模方法来给 Intel 公司股票的月对数收益率建立一个简单的 ARCH 模型. 图 3-1 中收益率平方的 ACF 和 PACF 清楚地表明条件异方差性的存在. 因此，不必再进行任何统计检验来确认建立 ARCH 模型的必要性，我们要做的是识别 ARCH 模型的阶. 图 3-1 右下方的样本 PACF 启示我们 ARCH(3)可能是合适的，因此下面将对 Intel 股票的月对数收益率具体建立一个如下形式的模型:

$$r_t = \mu + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \alpha_3 a_{t-3}^2$$

假定 ε_t 是独立同分布的标准正态序列，我们得到的拟合模型为

$$r_t = 0.0196 + a_t, \quad \sigma_t^2 = 0.0090 + 0.2973 a_{t-1}^2 + 0.0900 a_{t-2}^2 + 0.0626 a_{t-3}^2$$

各个参数估计值的标准差分别是 0.006 2，0.001 3，0.088 7，0.064 5 和 0.077 7. 估计值满足 ARCH(3)模型的一般条件，其中 α_2 和 α_3 的估计值在 5%的水平下不是统计显著的，因此模型

可简化.

去掉两个不显著的参数，我们得到的模型为

$$r_t = 0.0213 + a_t,\quad \sigma_t^2 = 0.00998 + 0.4437 a_{t-1}^2 \tag{3.10}$$

其中各参数估计的标准差为 0.0062，0.00124 和 0.0938，且所有估计都是高度显著的. 图 3-4 显示的是标准化的抖动和标准化抖动的某些函数的样本 ACF. 标准化抖动$\{\tilde{a}_t\}$的 Ljung-Box 统计量的观测值 $Q(10)=12.53$，p 值为 0.25. $\{\tilde{a}_t^2\}$的 Ljung-Box 统计量的观测值 $Q(10)=17.23$，p 值为 0.07. 因此，在 5%的水平下，式(3.10)的 ARCH(1)模型对给定的数据是适合的.

90

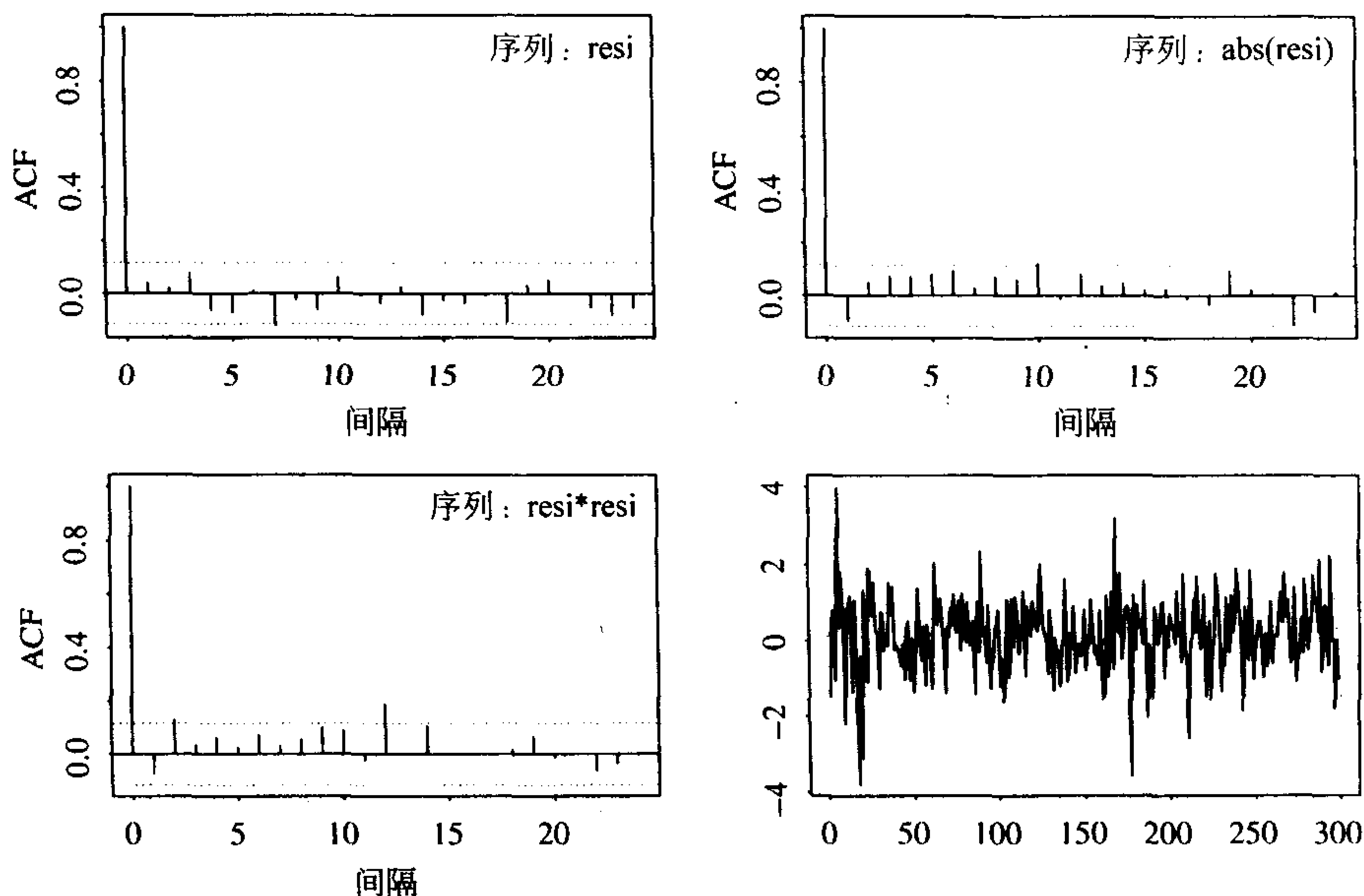


图 3-4 对 Intel 公司股票从 1973 年 1 月至 1997 年 12 月的月对数收益率建立高斯 ARCH(1)模型式(3.10)后，模型检验统计量：a)、b)、c)分别是标准化抖动本身、它们的平方和绝对值序列的样本 ACF，d)是标准化抖动的时间图(右下角)

式(3.10)的 ARCH(1)模型有一些有趣的性质：首先，Intel 公司股票的月对数收益率的期望值大约 2.1%，这是突出的；第二，$\hat{\alpha}_1^2=0.444^2<1/3$，从而月对数收益率的无条件四阶矩是有限的；第三，$r_t$ 的无条件方差是 $0.00998/(1-0.4437)=0.0179$. 综合起来，该 ARCH(1)模型可以用来预测 Intel 股票收益率的月波动率.

***t* 分布的新息**

为了比较，在 ϵ_t 服从标准化的自由度为 5 的学生-t 分布的假定下，给该序列拟合一个 ARCH(1)模型.

91

$$r_t = 0.0222 + a_t,\quad \sigma_t^2 = 0.0121 + 0.3029 a_{t-1}^2 \tag{3.11}$$

其中各参数的标准差分别为 0.0061，0.0019 和 0.1443. 所有估计值在 5%的水平下是显著的，

但$\hat{\alpha}_1$ 的 t 比仅为 2.10. a_t 的无条件方差是 $0.0121/(1-0.3029)=0.0174$，这与正态性假定下得到的 t 比值相近. 标准化扰动的 Ljung-Box 统计量的观测值为 $Q(10)=13.66$，p 值为 0.19，证实均值方程是合适的. 但是，标准化扰动平方的 Ljung-Box 统计量的观测值为 $Q(10)=23.83$，p 值为 0.008，波动率方程在 5%的水平下是不合适的. 我们改进此模型，考虑 ARCH(2)模型，得到

$$r_t = 0.0225 + a_t, \quad \sigma_t^2 = 0.0113 + 0.226a_{t-1}^2 + 0.108a_{t-2}^2 \tag{3.12}$$

其中参数的标准差分别为 0.006，0.002，0.135 和 0.094. a_{t-1}^2的系数在 10%水平下是显著的，但 a_{t-2}^2的系数仅仅比其标准差稍微大一点. 这时标准化扰动平方的 Ljung-Box 统计量的值为 $Q(10)=8.82$，p 值为 0.55. 因此，所拟合的 ARCH(2)模型是合适的.

比较式(3.10)、式(3.11)和式(3.12)三个模型，我们可以看到以下几点：(a)使用具有厚尾分布的 ϵ_t 可以降低 ARCH 效应；(b)对这个特殊数据集来说，这三个模型的差别是不大的. 对这组数据更适合的条件异方差模型是 GARCH(1，1)模型，这将在下一节讨论.

例 3.2 考虑马克对美元十分钟时间间隔汇率的百分比变化率. 图 3-2a 已经展示了该数据. 如图 3-3a 显示的，此序列没前后相关关系，但其平方的样本 PACF 有些大的值，特别是间隔为 1 和 3 时. 在某些较大的间隔时，PACF 有较大的值，但较小的间隔更重要一些. 按照前一小节讨论的方法，我们将对该序列用 ARCH(3)模型. 利用条件高斯似然函数，我们得到拟合的模型：

$$\sigma_t^2 = 0.22\times10^{-6} + 0.328a_{t-1}^2 + 0.073a_{t-2}^2 + 0.103a_{t-3}^2$$

其中所有的估计在 5%的水平下是统计显著的，参数的标准差分别是 0.46×10^{-8}，0.0162，0.0160 和 0.0147，用标准化的扰动 $\tilde{a}_t$ 进行模型检验，显示该模型是合适的.

注释：这章中条件异方差模型的估计是用 RATS(时间序列的回归分析)软件包进行的. 也有其他的软件可用，如 Eviews，SCA(Scientific Computing Associates，科学计算助手)和 S-Plus.

92

3.4 GARCH 模型

虽然 ARCH 模型简单，但为了充分地描述资产收益率的波动率过程，往往需要许多参数. 例如，标准普尔 500 指数的月超出收益率，其波动率过程需要 ARCH(9)模型来刻画. 这样就必须寻找其他模型. Bollerslev 在 1986 年提出了一个有用的推广形式，称为推广的 ARCH (GARCH)模型. 对对数收益率序列 r_t，我们假定其均值方程是一个 ARMA 模型. 设 $a_t=r_t-\mu_t$ 是均值修正的对数收益率，若

$$a_t = \sigma_t\epsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m}\alpha_i a_{t-i}^2 + \sum_{j=1}^{s}\beta_j\sigma_{t-j}^2 \tag{3.13}$$

其中$\{\epsilon_t\}$是一个独立同分布的随机变量序列，均值为 0，方差为 1，$\alpha_0>0$，$\alpha_i\geqslant0$，$\beta_j\geqslant0$，$\sum_{i=1}^{\max(m,s)}(\alpha_i+\beta_i)<1$ (这里对 $i>m$，$\alpha_i=0$，对 $j>s$，$\beta_j=0$)，则称 a_t 服从 GARCH$(m，s)$模型.

对 $\alpha_i+\beta_i$ 的限制条件保证 a_t 的无条件方差是有限的，同时它的条件方差 σ_t^2 是随时间变化的. 如前面一样，通常假定 ϵ_t 是标准正态的或标准化的学生-t 分布. 若 $s=0$，方程(3.13)就简化成一个纯 ARCH(m)模型.

为了弄清 GARCH 模型的性质，采用如下表示是有益的：令 $\eta_t=a_t^2-\sigma_t^2$，即 $\sigma_t^2=a_t^2-\eta_t$，把 $\sigma_{t-i}^2=a_{t-i}^2-\eta_{t-i}(i=0, \cdots, s)$带入式(3.13)，我们就能把 GARCH 模型写成如下形式

$$a_t^2=\alpha_0+\sum_{i=1}^{\max(m,s)}(\alpha_i+\beta_i)a_{t-i}^2+\eta_t-\sum_{j=1}^{s}\beta_j\eta_{t-j} \tag{3.14}$$

容易验证$\{\eta_t\}$是一个鞅差序列(即 $E\eta_t=0$ 且对 $j\geqslant 1$ 有 $\mathrm{Cov}(\eta_t, \eta_{t-j})=0$. 但$\{\eta_t\}$一般不是独立同分布序列. 式(3.14)对平方序列 a_t^2 来说是 ARMA 形式，因此 GARCH 模型可认为是 ARMA 的思想对平方序列 a_t^2 的应用. 利用 ARMA 模型的无条件均值，我们有

$$E(a_t^2)=\frac{\alpha_0}{1-\sum_{i=1}^{\max(m,s)}(\alpha_i+\beta_i)},$$

只要上式的分母是正的.

GARCH 模型的优点和缺点可以通过对最简单的 GARCH(1, 1)模型的分析看出. [93] GARCH(1, 1)模型为

$$\sigma_t^2=\alpha_0+\alpha_1 a_{t-1}^2+\beta_1\sigma_{t-1}^2,\quad 0\leqslant\alpha_1,\quad \beta_1\leqslant 1,\quad \alpha_1+\beta_1<1 \tag{3.15}$$

首先，大的 a_{t-1}^2或 σ_{t-1}^2引起大的 σ_t^2，这意味着大的 a_{t-1}^2会紧跟着另一个大的 a_t^2，这样就会产生在金融时间序列中有名的波动率聚类现象. 第二，可以证明：若$1-2\alpha_1^2-(\alpha_1+\beta_1)^2>0$，则

$$\frac{E(a_t^4)}{[E(a_t^2)]^2}=\frac{3[1-(\alpha_1+\beta_1)^2]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}>3$$

从而，与 ARCH 模型类似，GARCH(1, 1)过程的分布尾部比正态分布尾部厚. 第三，此模型提供了一个可用于描述波动率演变的简单的参数函数.

GARCH 模型的预测可用类似于 ARMA 模型的方法得到. 考虑式(3.15)的 GARCH(1, 1)模型，假定 h 为预测原点. 对向前 1 步预测，我们有

$$\sigma_{h+1}^2=\alpha_0+\alpha_1 a_h^2+\beta_1\sigma_h^2,$$

其中 a_h 和σ_h^2 是在时间指标为 h 时已知的. 因此，向前 1 步预测为

$$\sigma_h^2(1)=\alpha_0+\alpha_1 a_h^2+\beta_1\sigma_h^2$$

对向前多步预测，我们用 $a_t^2=\sigma_t^2\epsilon_t^2$，并把式(3.15)的波动率方程改写成

$$\sigma_{t+1}^2=\alpha_0+(\alpha_1+\beta_1)\sigma_t^2+\alpha_1\sigma_t^2(\epsilon_t^2-1)$$

当 $t=h+1$ 时，此方程变为

$$\sigma_{h+2}^2=\alpha_0+(\alpha_1+\beta_1)\sigma_{h+1}^2+\alpha_1\sigma_{h+1}^2(\epsilon_{h+1}^2-1)$$

因为 $E(\epsilon_{h+1}^2-1\mid F_h)=0$，故以 h 为预测原点的波动率的向前 2 步预测满足

$$\sigma_h^2(2)=\alpha_0+(\alpha_1+\beta_1)\sigma_h^2(1)$$

一般地，我们有

$$\sigma_h^2(l) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(l-1), \qquad l > 1 \tag{3.16}$$

这个结果与自回归多项式为 $1-(\alpha_1+\beta_1)B$ 的 ARMA(1，1)模型相同. 对式(3.16)重复代入，我们得到向前 l 步预测，写成 [94]

$$\sigma_h^2(l) = \frac{\alpha_0[1-(\alpha_1+\beta_1)^{l-1}]}{1-\alpha_1-\beta_1} + (\alpha_1+\beta_1)^{l-1}\sigma_h^2(1)$$

从而，只要 $\alpha_1+\beta_1<1$，就有

$$\sigma_h^2(l) \to \frac{\alpha_0}{1-\alpha_1-\beta_1}, \qquad \text{当 } l\to\infty \text{ 时.}$$

因此，只要 a_t 的无条件方差 $\mathrm{Var}(a_t)$存在，GARCH(1，1)模型的向前多步的波动率预测当预测步长趋于无穷时是收敛于 $\mathrm{Var}(a_t)$的.

关于 GARCH 模型的文献不是非常多，参见 Bollerslev，Chou and Kroner(1992)，Bollerslev，Engle 和 Nelson(1994)，以及这两篇文章的参考文献. GARCH 模型与 ARCH 模型有相同的弱点，例如，它对正的和负的抖动有相同的反应. 另外，新近关于高频金融时间序列的经验研究显示 GARCH 模型的尾部太薄，即使新息是服从学生-t 分布的 GARCH 模型，也不足以描述实际高频数据的尾部.

3.4.1 一个例子

ARCH 模型的建模方法也可以用来建立一个 GARCH 模型. 然而，GARCH 模型的阶不太容易确定. 在实际应用中，只用到低阶的 GARCH 模型，如 GARCH(1，1)、GARCH(2，1)、GARCH(1，2)等. 假定波动率$\{\sigma_t^2\}$的初始值是已知的，条件最大似然法仍然可以用. 例如，考虑 GARCH(1，1)模型，若 σ_1^2 已知固定，则 σ_t^2 可以递推得到. 在一些应用中，a_t 的样本方差作为 σ_1^2 的初始值较好. 可以用标准化的残差 $\tilde{a}_t=a_t/\sigma_t$ 和它的平方来检验所拟合的模型的合适性.

例 3.3 我们考虑标准普尔 500 指数月超出收益率，从 1926 年开始，共 792 个观察值序列如图 3-5 所示. 记 r_t 为超出收益率序列，图 3-6 所示的是 r_t 的样本 ACF 和 r_t^2 的样本 PACF. r_t 序列在间隔为 1 和间隔为 3 时有少许前后相关性，但主要特征还是 r_t^2 的 PACF 显示强烈的线性相关性. 若拟合 MA(3)模型，则得到

$$r_t = 0.006\,2 + a_t + 0.094\,4a_{t-1} - 0.140\,7a_{t-3}, \quad \hat{\sigma}_a = 0.057\,6,$$

对于序列，其中所有的系数在 5%的水平下都是显著的. 然而，为简单起见，我们使用 AR(3)模型：

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \beta_0 + a_t.$$ [95]

在正态性的假定下，所拟合的 AR(3)模型为

$$r_t = 0.088r_{t-1} - 0.023r_{t-2} - 0.123r_{t-3} + 0.006\,6 + a_t, \quad \hat{\sigma}_a^2 = 0.003\,33. \tag{3.17}$$

对 GARCH 效应，我们用 GARCH(1，1)模型

$$a_t = \sigma_t\epsilon_t, \quad \sigma_t^2 = \alpha_0 + \beta_1\sigma_{t-1}^2 + \alpha_1 a_{t-1}^2.$$

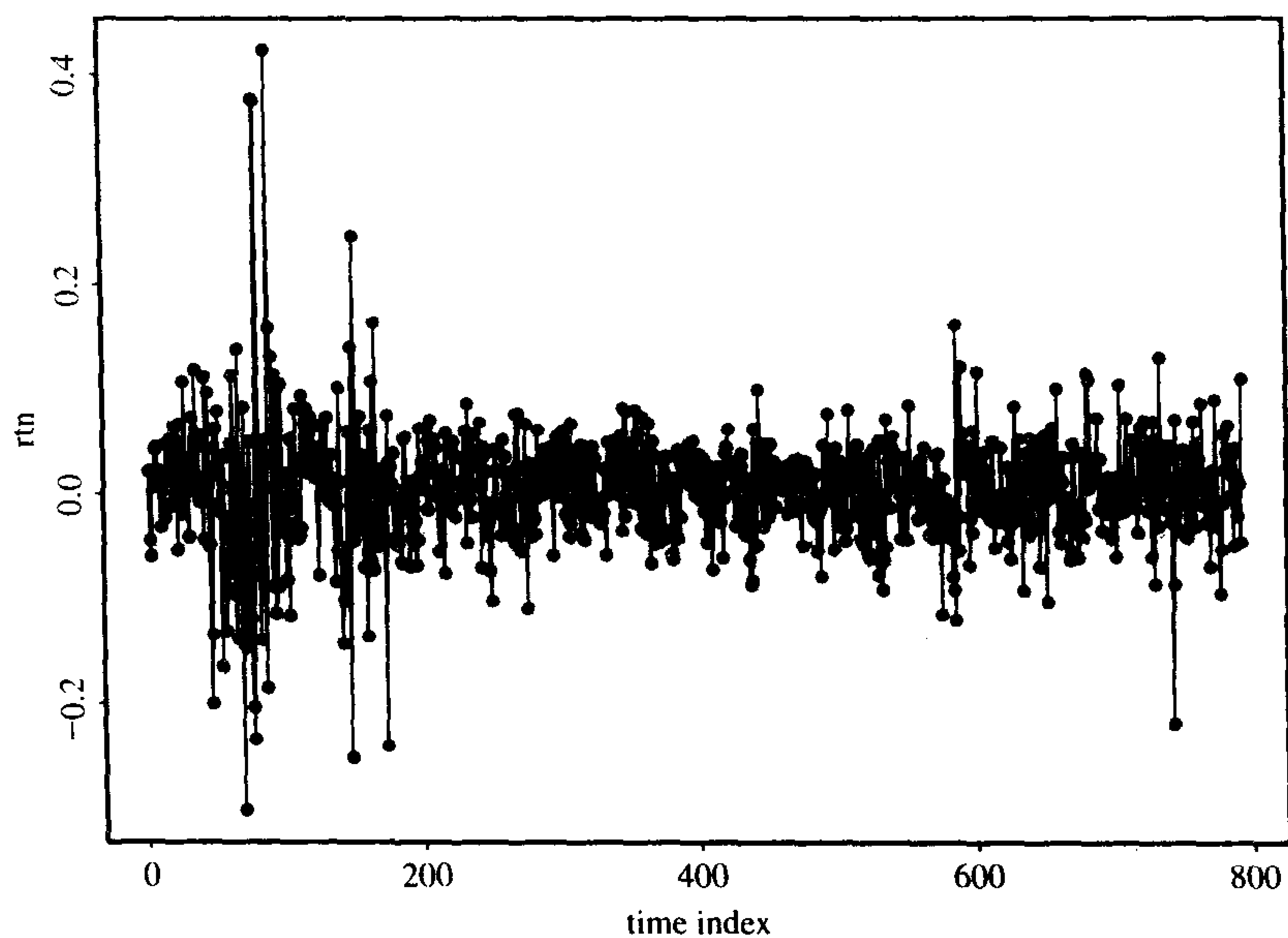
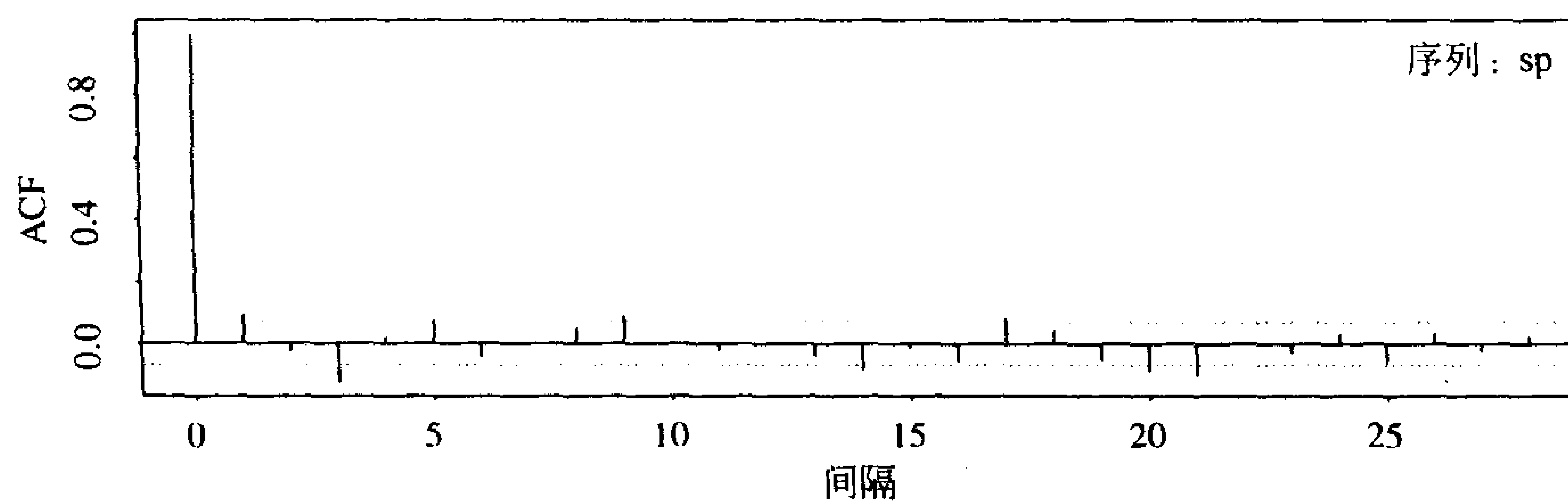
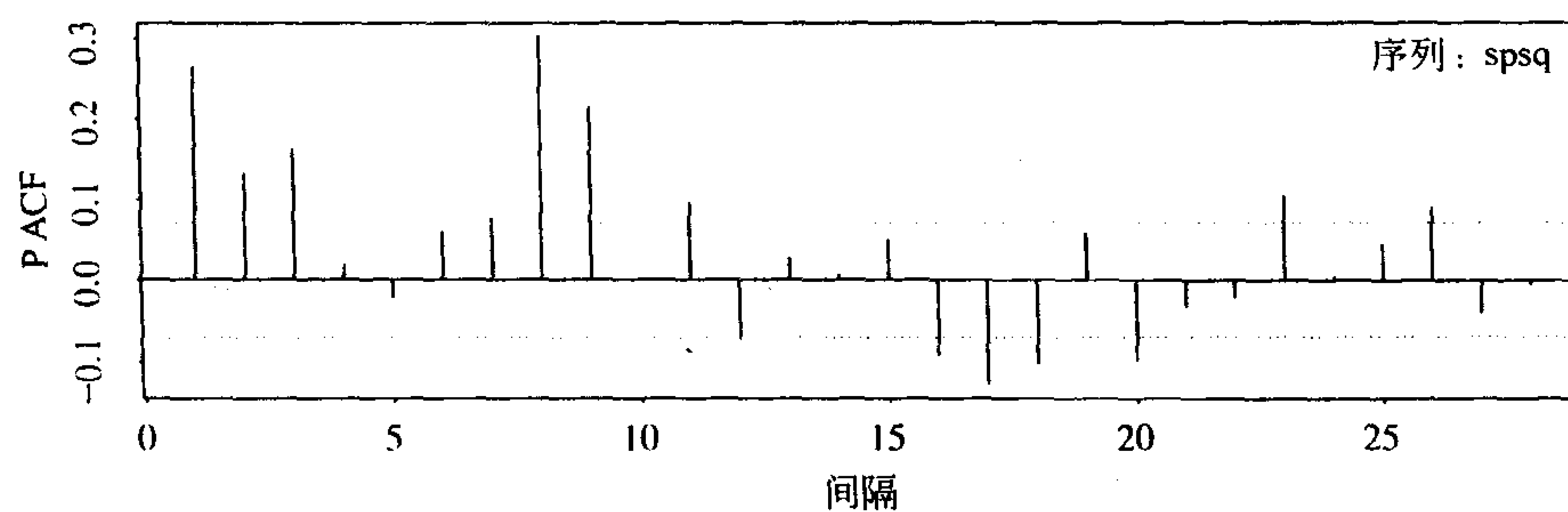


图 3-5 标准普尔 500 指数月超出收益率的时间序列图



a) 标准普尔500指数月超出收益率的样本自相关函数



b) 月超出收益率平方的样本偏自相关函数(PACF)

图 3-6

对该 AR(3)-GARCH(1，1)模型的联合估计得到

$$r_t = 0.021r_{t-1} - 0.034r_{t-2} - 0.013r_{t-3} + 0.0085 + a_t,$$

$$\sigma_t^2 = 0.000099 + 0.8476\sigma_{t-1}^2 + 0.1219a_{t-1}^2.$$

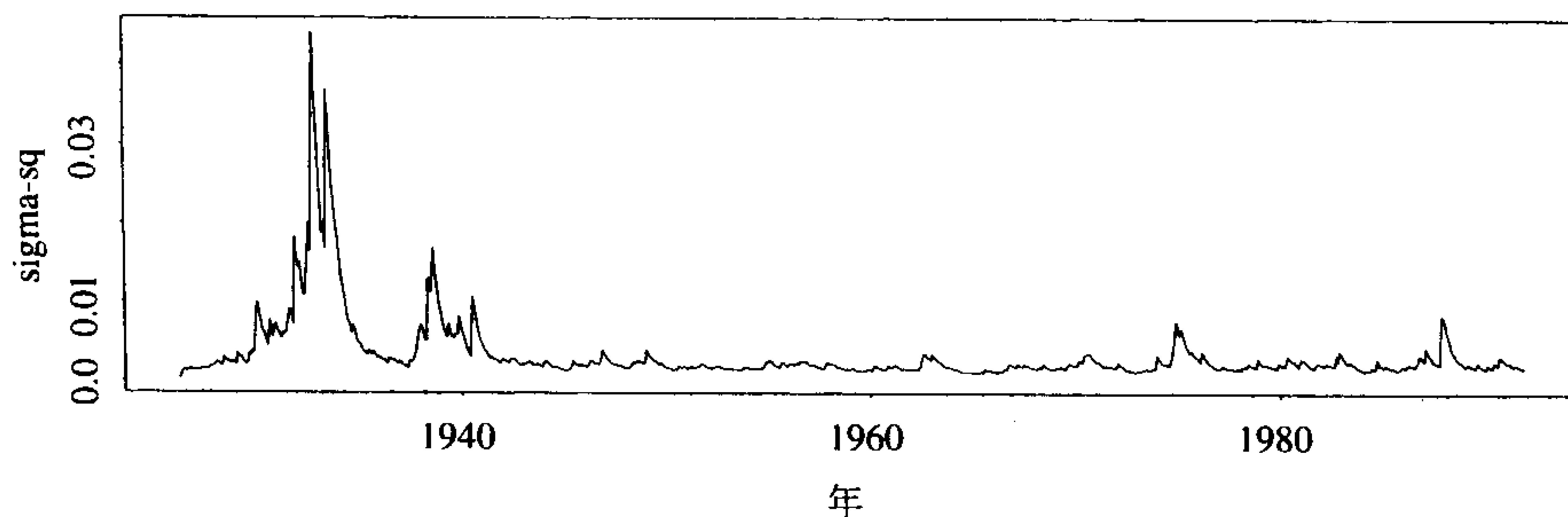
由波动率方程，a_t 的无条件方差为

$$\frac{0.000099}{1-0.8476-0.1219} = 0.00325,$$

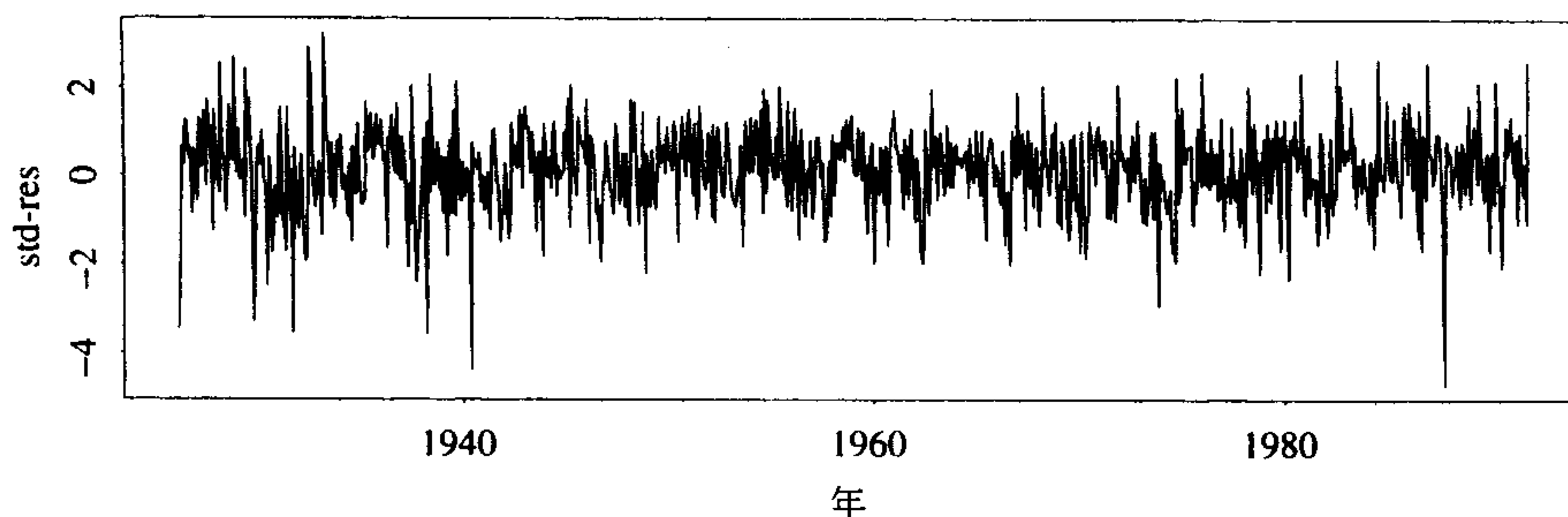
这与式(3.17)中$\hat{\sigma}_a^2$ 非常接近. 然而，均值方程中的各参数的 t 比表明自回归系数在 5%的水平下都是不显著. 因此，我们撤掉所有自回归系数来改进模型，所拟合的改进模型为

$$r_t = 0.0065 + a_t, \quad \sigma_t^2 = 0.00014 + 0.8220\sigma_{t-1}^2 + 0.1352a_{t-1}^2. \tag{3.18}$$

均值方程中参数的标准差为 0.0015，而波动率方程中的三个参数的标准差分别为 0.00002、0.0208 和 0.0166. a_t 的无条件方差为 0.0001/(1−0.822−0.1352)=0.00324. 这是一个简单的、平稳的 GARCH(1，1)模型. 图 3-7 所示的是所估计出的式(3.18)的 GARCH(1，1)模型波动率过程和标准化的抖动$\tilde{a}_t = a_t/\sigma_t$. $\tilde{a}_t$ 序列看起来像白噪声过程，图 3-8 提供了标准化抖动$\tilde{a}_t$ 和抖动平方$\tilde{a}_t^2$ 的样本自相关函数，这两个 ACF 显示在这两个过程中没有显著的前后相关性. 更具体地说，对$\tilde{a}_t$，我们有 $Q(10)=10.32(0.41)$和 $Q(20)=22.66(0.31)$，对$\tilde{a}_t^2$，有 $Q(10)=8.83(0.55)$，$Q(20)=15.82(0.73)$，括号中的数为对应检验统计量的 p 值. 因此，模型是合适的. 注意到



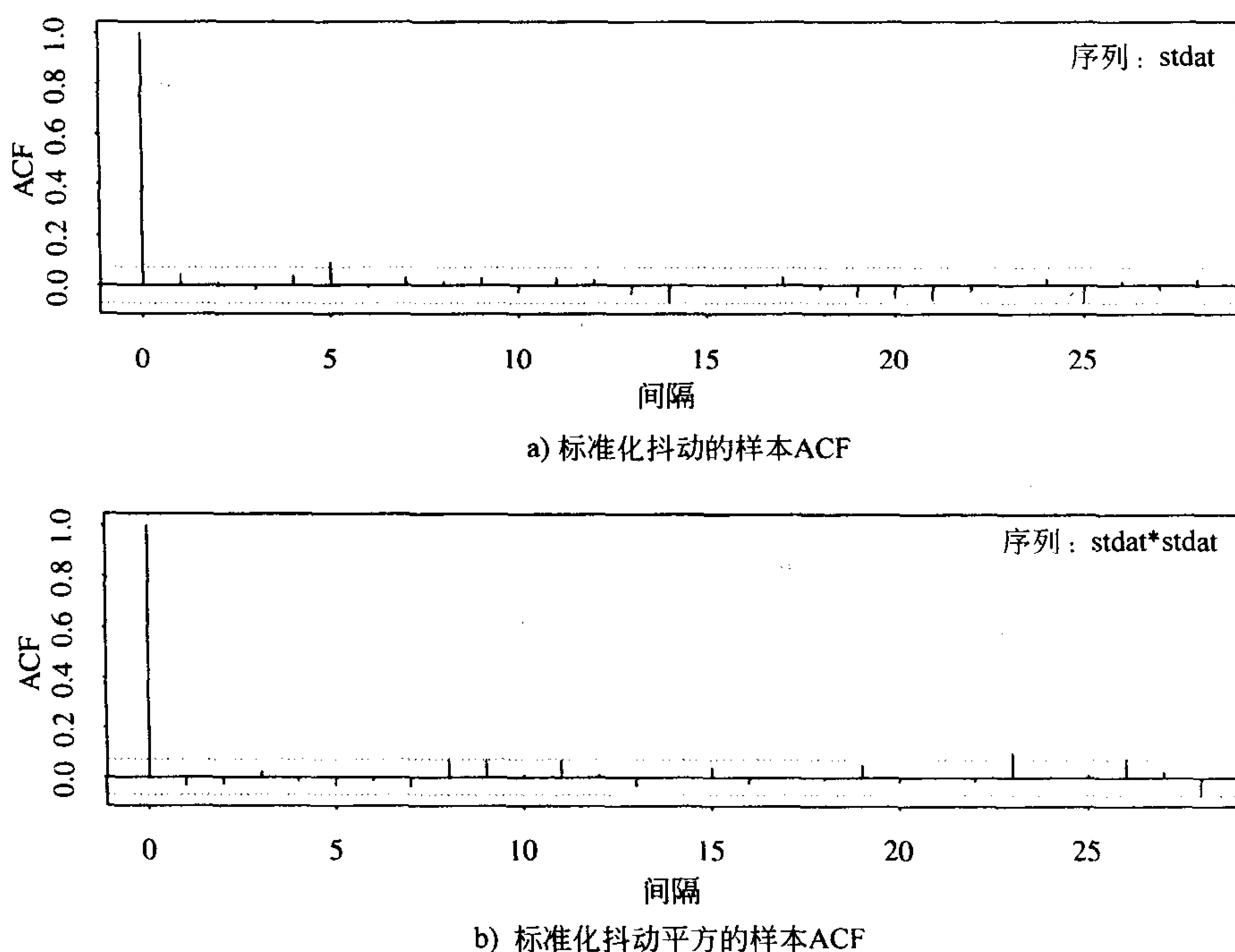
a) 对标准普尔500指数的月超出收益率估计出的波动率的时间序列图



b) 标准普尔500指数的月超出收益率的标准化抖动的时间序列图

图 3-7 两个图都是基于式(3.18)的 GARCH(1，1)模型

所拟合的模型显示$\hat{\alpha}_1+\hat{\beta}_1=0.9572$，非常接近于 1. 这种现象在实际中是常见的，以至于经常在 GARCH(1，1)模型中加上限制$\alpha_1+\beta_1=1$，从而引出求和 GARCH(或 IGARCH)模型，见第 3.5 节.



a) 标准化抖动的样本ACF

b) 标准化抖动平方的样本ACF

图 3-8 标准普尔 500 指数的月超出收益率所建的 GARCH(1，1)模型式(3.18)的模型检验

为了预测标准普尔 500 指数的月超出收益率的波动率，我们可以用式(3.18)中的波动率方程. 例如，在预测原点 h，有 $\sigma_{h+1}^2=0.00014+0.822\sigma_h^2+0.1352a_h^2$. 向前 1 步预测是

$$\sigma_h^2(1)=0.00014+0.822\sigma_h^2+0.1352a_h^2$$

其中 a_h 是 t 时刻均值方程的残差，σ_h^2 由波动率方程得到. 初始值 σ_0^2 可取 0 或 a_t 的无条件方差. 对多步向前预测，我们利用式(3.16)递推. 表 3-1 所示的是基于式(3.18)的 GARCH(1，1)模型、以 $h=792$ 为预测原点对标准普尔 500 指数的月超出收益率的部分均值和波动率预测的结果.

表 3-1

预测步长	1	2	3	4	5	∞
收益率	0.006 5	0.006 5	0.006 5	0.006 5	0.006 5	0.006 5
波动率	0.003 11	0.003 12	0.003 12	0.003 13	0.003 14	0.003 24

注：标准普尔 500 指数的月超出收益率的波动率预测. 预测原点为 $h=792$，对应的时间为 1991 年 12 月. 这里，波动率是条件方差.

带服从 t 分布的新息

假定 ϵ_t 服从自由度为 5 的标准化学生-t 分布，我们重新估计 GARCH(1，1)模型，得到

$$r_t = 0.008\,5 + a_t, \quad \sigma_t^2 = 0.000\,18 + 0.127\,2a_{t-1}^2 + 0.821\,7\sigma_{t-1}^2 \tag{3.19}$$

其中参数的标准差分别为 0.001 4、0.55×10^{-4}、0.034 9 和 0.038 2，这个模型本质上是一个 IGARCH(1，1)模型，因为 $\hat{\alpha}_1+\hat{\beta}_1\approx0.96$，近似于 1. 标准化残差的 Ljung-Box 统计量 $Q(10)=10.45$，p 值为 0.40，而$\{\tilde{a}_t^2\}$序列的 Ljung-Box 统计量 $Q(10)=9.33$，p 值为 0.50. 这样所拟合的带学生-t 分布的 GARCH(1，1)模型是合适的.

自由度的估计

如果我们更进一步地拓展上述 GARCH(1，1)模型，其中学生-t 分布的自由度通过数据估计，得到的模型

$$r_t = 0.008\,3 + a_t, \quad \sigma_t^2 = 0.000\,17 + 0.122\,7a_{t-1}^2 + 0.819\,3\sigma_{t-1}^2 \tag{3.20}$$

其中所估计的自由度为 6.51. 式(3.20)中参数估计值的标准差与式(3.19)非常接近，自由度估计的标准差为 1.49. 从而，在 5%的水平下我们不能拒绝使用自由度为 5 的标准化学生-t 分布的假设.

3.4.2 预测的评价

由于资产收益率的波动率是不能直接观测的，比较不同波动率模型的预测表现对数据分析师来说是一个挑战. 文献中，有的研究人员利用样本以外(out-of-sample)的预测，比较波动率的预测值 $\sigma_h^2(l)$与预测样本中的抖动 a_{h+l}^2来评价波动率模型的预测表现.（也就是把整个样本分成两个部分，用一部分(往往是大部分)来预测另一部分(称为预测样本)，可得 a_{h+l} 和 $\sigma_h^2(l)$. ——译者注)用这个方法常常会发现 a_{h+l}^2和 $\sigma_h^2(l)$有较低的相关系数. 这个发现是不奇怪的，因为单独 a_{h+l}^2本身不能作为 $h+l$ 时刻波动率的合适度量. 例如，对向前 1 步预测，从统计观点讲，$E(a_{h+1}^2\mid F_h)=\sigma_{h+1}^2$，因此 a_{h+1}^2是 σ_{h+1}^2的相合估计. 但它不是 σ_{h+1}^2的精确估计，因为一个有已知均值的随机变量的单个观察不可能提供该随机变量方差的精确估计. 从而，严格地说，这种评价波动率模型的预测表现的方法是不合适的. 关于 GARCH 模型的预测评价的更多信息，读者可参见 Andersen 和 Bollerslev(1998).

3.5 求和 GARCH 模型

若式(3.14)中的 GARCH 表示的 AR 多项式有一个单位根，则我们就有一个 IGARCH 模型. 因此，IGARCH 模型就是单位根 GARCH 模型. 类似于 ARIMA 模型，IGARCH 模型主要特点是过去抖动的平方 $\eta_{t-i}=a_{t-i}^2-\sigma_{t-i}^2(i>0)$对 a_t^2 的影响是持久的.

IGARCH(1，1)模型能写成：

$$a_t = \sigma_t\epsilon_t, \quad \sigma_t^2 = \alpha_0 + \beta_1\sigma_{t-1}^2 + (1-\beta_1)a_{t-1}^2,$$

其中$\{\epsilon_t\}$与前面一样定义，$1>\beta_1>0$. 用标准普尔 500 指数的每月超出收益率数据估计出的 IGARCH(1.1)模型为

[96 ~ 100]

$$r_t = 0.0067 + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 0.000119 + 0.8059\sigma_{t-1}^2 + 0.1941a_{t-1}^2$$

其中波动率方程中各估计的标准差依次是 0.001 7、0.000 013、0.014 4. 参数的估计值与前面的 GARCH(1, 1)模型很接近，但这两个模型间有一个大的差别. 在 IGARCH(1, 1)模型下，a_t 的无条件方差，从而 r_t 的无条件方差，是没有定义的. 这一点对超出收益率序列是难以验证的. 从理论观点看，IGARCH 现象可能会是波动率常有的水平移动所引起的. 波动率持续性的真正原因值得仔细研究.

当 $\alpha_1+\beta_1=1$ 时，由式(3.16)重复代入

$$\sigma_h^2(l) = \sigma_h^2(1) + (l-1)\alpha_0, \qquad l \geqslant 1,$$

其中 h 是预测原点. 由此，$\sigma_h^2(1)$对将来波动率的效应也是持续的，波动率预测形成了一个斜率为 α_0 的直线. Nelson(1990)研究了在 IGARCH 模型下波动率过程 σ_t^2 的某些概率性质. 过程 σ_t^2 是一个鞅，对它在文献中有一些漂亮的结果可用. 在一些条件下，波动率过程是严平稳的，但不是弱平稳的. 因为它的头两阶矩不存在.

在研究 IGARCH(1, 1)模型时，$\alpha_0=0$ 的情形是特别令人感兴趣的. 这时，对所有预测步长，波动率的预测值都是 $\sigma_h^2(1)$. 这个特殊的 IGARCH(1, 1)模型正是风险度量系统 RiskMetrics 所用的波动率模型，这个系统是一种计算风险值(Value at Risk)的方法(见第 7 章).

3.6 GARCH-M 模型

在金融中，证券的收益率会依赖于它的波动率. 为了对这种现象建模，人们会考虑 GARCH-M 模型，其中“M”表示均值中的 GARCH(GARCH in mean). 简单的 GARCH(1, 1)-M 模型能写成

$$r_t = \mu + c\sigma_t^2 + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3.21}$$

其中 μ 和 c 是常数. 参数 c 叫做风险溢价参数. c 为正值意味着收益率与它过去的波动率成正相关. 文献中还出现过其他的风险溢价的具体化，如 $r_t=\mu+c\sigma_t+a_t$.

式(3.21)的 GARCH-M 模型蕴涵了收益率序列 r_t 存在前后相关性，这种前后相关性是由波动率过程 $\{\sigma_t^2\}$的前后相关性导致的. 因此，风险溢价的存在是历史股票收益率具有前后相关性的另一种原因.

为了说明 GARCH-M 模型的应用，我们考虑给标准普尔 500 指数从 1926 年 1 月至 1991 年 12 月的月超出收益率拟合一个 GARCH(1, 1)-M 模型. 所拟合的模型为

$$r_t = 0.0028 + 1.99\sigma_t^2 + a_t,$$
$$\sigma_t^2 = 0.00016 + 0.1328a_{t-1}^2 + 0.8137\sigma_{t-1}^2$$

[101] 其中均值方程中两个参数的标准差分别为 0.002 2 和 0.742 5，波动率方程中三个参数的标准差分别是 0.000 02、0.022 0 和 0.019 2. 该指数收益率的风险溢价的估计值是正的，并且在 5%

的水平下是显著的. 风险溢价的思想也可应用于其他 GARCH 模型.

3.7 指数 GARCH 模型

为了克服 GARCH 模型在处理金融时间序列时的一些弱点，Nelson(1991)提出了指数 GARCH(EGARCH)模型. 具体地，为了允许在模型中体现正的和负的资产收益率的非对称效应，他考虑加权的新息：

$$g(\epsilon_t)=\theta\epsilon_t+\gamma[|\epsilon_t|-E(|\epsilon_t|)], \tag{3.22}$$

其中 θ 和 γ 是实常数. ϵ_t 和 $|\epsilon_t|-E(|\epsilon_t|)$ 都是零均值的 iid 序列，有连续的分布. 因此，$E[g(\epsilon_t)]=0$. $g(\epsilon_t)$ 的非对称性容易从下式看出：

$$g(\epsilon_t)=\begin{cases}(\theta+\gamma)\epsilon_t-\gamma E(|\epsilon_t|), & \text{如果 } \epsilon_t\geqslant 0\\ (\theta-\gamma)\epsilon_t-\gamma E(|\epsilon_t|), & \text{如果 } \epsilon_t<0\end{cases}$$

注释：对标准高斯随机变量 ϵ_t，$E(|\epsilon_t|)=\sqrt{2/\pi}$. 对式(3.7)中的标准化学生-$t$ 分布，我们有

$$E(|\epsilon_t|)=\frac{2\sqrt{v-2}\,\Gamma((v+1)/2)}{(v-1)\Gamma(v/2)\sqrt{\pi}}$$

EGARCH(m, s)模型能写成：

$$a_t=\sigma_t\epsilon_t,\quad \ln(\sigma_t^2)=\alpha_0+\frac{1+\beta_1B+\cdots+\beta_sB^s}{1-\alpha_1B-\cdots-\alpha_mB^m}g(\epsilon_{t-1}) \tag{3.23}$$

其中 α_0 是常数，B 是向后推移算子使得 $Bg(\epsilon_t)=g(\epsilon_{t-1})$，$1+\beta_1B+\cdots+\beta_sB^s$ 和 $1-\alpha_1B-\cdots-\alpha_mB^m$ 是多项式，它们的根都在单位圆外且没有公因子. 根在单位圆外，意指根的模是大于 1 的. 式(3.23)又一次用通常 ARMA 参数化形式来描述 a_t 的条件方差随时间的演变. 基于这个表示，EGARCH 模型的一些性质可以用与对 GARCH 模型所用的类似方法得到，例如 $\ln(\sigma_t^2)$ 的无条件均值为 α_0. 然而，这个模型在如下几点上与 GARCH 模型不同：第一，它使用条件方差的对数，放松了对模型系数非负性的限制；第二，$g(\epsilon_t)$ 的使用使模型对 a_t 的正的和负的延迟值，模型的反应不对称. EGARCH 模型的其他性质可在 Nelson(1991)中找到. [102]

为了更好地理解 EGARCH 模型，我们考虑阶为(1，0)时的简单模型：

$$a_t=\sigma_t\epsilon_t,\quad (1-\alpha B)\ln(\sigma_t^2)=(1-\alpha)\alpha_0+g(\epsilon_{t-1}) \tag{3.24}$$

其中 $\epsilon_t s$ 是 iid 标准正态序列，α_1 的下标被省略了. 在这种情形，$E(|\epsilon_t|)=\sqrt{2/\pi}$，$\ln(\sigma_t^2)$ 的模型变成

$$(1-\alpha B)\ln(\sigma_t^2)=\begin{cases}\alpha_*+(\theta+\gamma)\epsilon_{t-1}, & \text{当 } \epsilon_{t-1}\geqslant 0 \text{ 时}\\ \alpha_*+(\theta-\gamma)\epsilon_{t-1}, & \text{当 } \epsilon_{t-1}<0 \text{ 时}\end{cases} \tag{3.25}$$

其中 $\alpha_*=(1-\alpha)\alpha_0-\sqrt{2/\pi}\,\gamma$. 这是一个非线性函数，类似于 Tong(1978，1990)中的门限自回归(TAR，threshold autoregressive)模型. 这里只要知道：对这个简单的 EGARCH 模型来说，条件方差以非线性方式依赖于 a_{t-1} 的符号. 具体地说，我们有

$$\sigma_t^2 = \sigma_{t-1}^{2\alpha}\exp(\alpha_*)\begin{cases}\exp\left[(\theta+\gamma)\dfrac{a_{t-1}}{\sqrt{\sigma_{t-1}^2}}\right], & \text{当 } a_{t-1} \geqslant 0 \text{ 时}, \\ \exp\left[(\theta-\gamma)\dfrac{a_{t-1}}{\sqrt{\sigma_{t-1}^2}}\right], & \text{当 } a_{t-1} < 0 \text{ 时}, \end{cases}$$

系数$(\theta+\gamma)$和$(\theta-\gamma)$表明对正的和负的a_{t-1}响应的非对称性．因此，当$\theta\neq0$时，模型是非线性的．对高阶 EGARCH 模型来讲，非线性变得复杂得多．Cao 和 Tsay(1992)利用非线性模型，包括 EGARCH 模型，得到向前多步波动率的预报．第 4 章中将讨论金融时间序列的非线性．

3.7.1 实例说明

Nelson(1991)把 EGARCH 模型应用到价值加权市场指数的月超出收益率上，数据是从证券价格研究中心得到的，从 1962 年 7 月至 1987 年 12 月．超出收益率是用价值加权指数的收益率减去国库券的月收益率，假定一个月中每一天的国库券收益率是常数．共有 6 408 个观察值．用r_t表示超出收益率，所用模型为

$$r_t = \phi_0 + \phi_1 r_{t-1} + c\sigma_t^2 + a_t \tag{3.26}$$

$$\ln(\sigma_t^2) = \alpha_0 + \ln(1 + wN_t) + \frac{1+\beta B}{1-\alpha_1 B - \alpha_2 B^2} g(\epsilon_{t-1})$$

其中σ_t^2是给定F_{t-1}下a_t的条件方差，N_t是第$t-1$个交易日和第t个交易日之间不交易的天数，α_0和w是实参数，$g(\epsilon_t)$由式(3.22)定义，ϵ_t服从广义误差分布，其概率密度函数是

103

$$f(x) = \frac{v\exp[-(1/2)|x/\lambda|^v]}{\lambda 2^{(1+1/v)}\Gamma(1/v)}, \quad -\infty < x < \infty, \quad 0 < v \leqslant \infty,$$

其中$\Gamma(.)$是 gamma 函数，

$$\lambda = [2^{(-2/v)}\Gamma(1/v)/\Gamma(3/v)]^{1/2}$$

与 GARCH-M 模型相似，式(3.26)中的参数c是风险溢价参数．表 3-2 给出了模型的参数估计值和它们的标准差．式(3.26)的模型均值方程有两条性质值得注意：第一，它用 AR(1)模型来刻画超出收益率中可能的前后相关性；第二，它用波动率σ_t^2作为回归变量来计算风险溢价．所估计出的风险溢价是负的，但是统计不显著的．

表 3-2

参数	α_0	w	γ	α_1	α_2	β
估计值	−10.06	0.183	0.156	1.929	−0.929	−0.978
标准差	0.346	0.028	0.013	0.015	0.015	0.006
参数	θ	ϕ_0	ϕ_1	c	v	
估计值	−0.118	3.5×10^{-4}	0.205	−3.361	1.576	
标准差	0.009	9.9×10^{-5}	0.012	2.026	0.032	

注：对价值加权 CRSP 市场指数的日超出收益率估计出的 AR(1)-EGARCH(2，1)模型，数据是从 1962 年 7 月至 1987 年 12 月．

3.7.2 另一个例子

我们来考虑 IBM 股票从 1926 年 1 月至 1997 年 12 月的月对数收益率，共 864 个观察值. 拟合一个 AR(1)-EGARCH(1, 0)模型：

$$r_t = 0.0105 + 0.092 r_{t-1} + a_t, \quad a_t = \sigma_t \varepsilon_t \tag{3.27}$$

$$\ln(\sigma_t^2) = -5.496 + \frac{g(\varepsilon_{t-1})}{1-0.856B},$$

$$g(\varepsilon_{t-1}) = -0.0795\varepsilon_{t-1} + 0.2647\left[|\varepsilon_{t-1}| - \sqrt{2/\pi}\right], \tag{3.28}$$

其中$\{\varepsilon_t\}$是一列独立的标准高斯随机变量. 所有参数估计值在 5%水平下都是统计显著的. 至于模型的检验，标准化的残值过程$\tilde{a}_t = a_t/\sigma_t$ 的 Ljung-Box 统计量为 $Q(10)=6.31(0.71)$和$Q(20)=21.4(0.32)$而对平方过程$\tilde{a}_t^2$，$Q(10)=4.13(0.90)$和 $Q(20)=15.93(0.66)$，括号中的参数是 p 值. 因此，在所拟合的标准残差序列中没有前后相关性或条件异方差性，说明所拟合的模型是合适的. [104]

由估计出的波动率方程(3.28)并利用$\sqrt{2/\pi} \approx 0.7979$，我们得到波动率方程为

$$\ln(\sigma_t^2) = -1.001 + 0.856\ln(\sigma_{t-1}^2) + \begin{cases} 0.1852\varepsilon_{t-1} & \text{当 } \varepsilon_{t-1} \geq 0 \text{ 时} \\ -0.3442\varepsilon_{t-1} & \text{当 } \varepsilon_{t-1} < 0 \text{ 时}. \end{cases}$$

做一个反对数变换，我们有

$$\sigma_t^2 = \sigma_{t-1}^{2\times 0.856} e^{-1.001} \times \begin{cases} e^{0.1852\varepsilon_{t-1}} & \text{当 } \varepsilon_{t-1} \geq 0 \text{ 时} \\ e^{-0.3442\varepsilon_{t-1}} & \text{当 } \varepsilon_{t-1} < 0 \text{ 时}. \end{cases}$$

此方程说明了在 EGARCH 模型下波动率对过去正的和负的抖动的不对称反应. 例如，当标准化的抖动抖动两个单位(即 2 个标准差)时，有

$$\frac{\sigma_t^2(\varepsilon_{t-1}=-2)}{\sigma_t^2(\varepsilon_{t-1}=2)} = \frac{\exp[-0.3442\times(-2)]}{\exp(0.1852\times 2)} = e^{0.318} = 1.374.$$

因此，变动两个标准差的负抖动对波动率的影响要比相同强度的正抖动的影响高 37.4%. 这个例子清楚地显示出 EGARCH 模型的不对称特征. 一般来说，抖动越大，正负抖动对波动率的影响的差别就越大.

3.7.3 用 EGARCH 模型预测

我们用 EGARCH(1, 0)模型来说明 EGARCH 模型的向前多步预测. 假定模型的参数已知，新息是标准高斯分布的，我们有

$$\ln(\sigma_t^2) = (1-\alpha_1)\alpha_0 + \alpha_1\ln(\sigma_{t-1}^2) + g(\varepsilon_{t-1}),$$

$$g(\varepsilon_{t-1}) = \theta\varepsilon_{t-1} + \gamma(|\varepsilon_{t-1}| - \sqrt{2/\pi})$$

两边取指数，模型变成

$$\sigma_t^2 = \sigma_{t-1}^{2\alpha_1}\exp[(1-\alpha_1)\alpha_0]\exp[g(\varepsilon_{t-1})],$$

$$g(\varepsilon_{t-1}) = \theta\varepsilon_{t-1} + \gamma(|\varepsilon_{t-1}| - \sqrt{2/\pi}). \tag{3.29}$$

设 h 是预测原点，对向前 1 步预测，我们有

$$\sigma_{h+1}^2 = \sigma_h^{2\alpha_1}\exp[(1-\alpha_1)\alpha_0]\exp[g(\epsilon_h)]$$

105 其中右边的所有量都是已知的. 因此以 h 为预测原点的向前 1 步的波动率预测是$\hat{\sigma}_h^2(1)=\sigma_{h+1}^2$，由上式给出. 对向前 2 步预测，式(3.29)给出

$$\sigma_{h+2}^2 = \sigma_{h+1}^{2\alpha_1}\exp[(1-\alpha_1)\alpha_0]\exp[g(\epsilon_{h+1})]$$

在 h 时刻取条件期望，我们有

$$\hat{\sigma}_h^2(2) = \hat{\sigma}_h^{2\alpha_1}(1)\exp[(1-\alpha_1)\alpha_0]E_h\{\exp[g(\epsilon_{h+1})]\},$$

其中 E_h 表示在时间原点 h 所取的条件期望. 上述期望可以由下式得到：

$$\begin{aligned}
E\{\exp[g(\epsilon)]\} &= \int_{-\infty}^{\infty}\exp\left[\theta\epsilon+\gamma(|\epsilon|-\sqrt{2/\pi})\right]f(\epsilon)\mathrm{d}\epsilon \\
&= \exp(-\gamma\sqrt{2/\pi})\left[\int_0^{\infty}\mathrm{e}^{(\theta+\gamma)\epsilon}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\epsilon^2/2}\mathrm{d}\epsilon\right. \\
&\quad \left.+\int_{-\infty}^{0}\mathrm{e}^{(\theta-\gamma)\epsilon}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\epsilon^2/2}\mathrm{d}\epsilon\right] \\
&= \exp(-\gamma\sqrt{2/\pi})\left[\mathrm{e}^{(\theta+\gamma)^2/2}\Phi(\theta+\gamma)+\mathrm{e}^{(\theta-\gamma)^2/2}\Phi(\gamma-\theta)\right]
\end{aligned}$$

其中 $f(\epsilon)$和 $\Phi(x)$分别表示标准正态分布的概率密度函数和累积分布函数. 因此，向前 2 步波动率预测值为

$$\begin{aligned}
\hat{\sigma}_h^2(2) = &\ \hat{\sigma}_h^{2\alpha_1}(1)\exp\left[(1-\alpha_1)\alpha_0-\gamma\sqrt{2/\pi}\right] \\
&\times[\exp\{(\theta+\gamma)^2/2\}\Phi(\theta+\gamma)+\exp\{(\theta-\gamma)^2/2\}\Phi(\gamma-\theta)]
\end{aligned}$$

重复前面的方法，我们对向前 j 步预测得到递推公式：

$$\begin{aligned}
\hat{\sigma}_h^2(j) = &\ \hat{\sigma}_h^{2\alpha_1}(j-1)\exp(\omega) \\
&\times[\exp\{(\theta+\gamma)^2/2\}\Phi(\theta+\gamma)+\exp\{(\theta-\gamma)^2/2\}\Phi(\gamma-\theta)]
\end{aligned}$$

其中 $\omega=(1-\alpha_1)\alpha_0-\gamma\sqrt{2/\pi}$. $\Phi(\theta+\gamma)$和 $\Phi(\gamma-\theta)$的值可以从统计软件包中得到. 另外，可用第 6 章附录 B 中的方法得到这些较精确的近似值.

作为例子，考虑前一小节为 IBM 股票月对数收益率建立的 AR(1)-EGARCH(1，0)模型. 利用所建立的 EGARCH(1，0)模型来计算该序列的波动率预测. 在预测原点 $t=864$，预测值

106 是$\hat{\sigma}_{864}(1)=6.05\times10^{-3}$，$\hat{\sigma}_{864}^2(2)=5.82\times10^{-3}$，$\hat{\sigma}_{864}^2(3)=5.63\times10^{-3}$，$\hat{\sigma}_{864}^2(10)=4.94\times10^{-3}$. 这些预测值逐步收敛到式(3.27)中的抖动过程的样本方差为 4.37×10^{-3}.

3.8 CHARMA 模型

文献中已经有很多经济计量模型被提出来描述式(3.2)中的条件方差 σ_t^2 的演变. 我们介绍一下条件异方差 ARMA(CHARMA)模型，此模型是用随机系数来产生条件异方差性(见 Tsay (1987)). CHARMA 模型与 ARCH 模型不同，但这两个模型有相同的二阶条件性质. CHARMA 模型定义为

$$r_t = \mu_t + a_t, \quad a_t = \delta_{1t}a_{t-1} + \delta_{2t}a_{t-2} + \cdots + \delta_{mt}a_{t-m} + \eta_t \tag{3.30}$$

其中$\{\eta_t\}$是均值为零、方差为σ_η^2的高斯白噪声序列．$\{\boldsymbol{\delta}_t\}=\{(\delta_{1t},\cdots,\delta_{mt})'\}$是均值为 0、非负定的协方差阵为$\boldsymbol{\Omega}$的独立同分布随机向量序列，并且$\{\boldsymbol{\delta}_t\}$与$\{\eta_t\}$独立．这节我们将用到向量与矩阵运算的一些基本性质来简化叙述．读者可以参考第 8 章的附录 A，对这些性质做一些简短的回顾．对$m>0$，模型可写成

$$a_t = \boldsymbol{a}'_{t-1}\boldsymbol{\delta}_t + \eta_t$$

其中$\boldsymbol{a}_{t-1}=(a_{t-1},\cdots,a_{t-m})'$是$a_t$的延迟值构成的向量，在$t-1$时刻是已知的．从而，式(3.30)的 CHARMA 模型中a_t的条件方差为

$$\begin{aligned}\sigma_t^2 &= \sigma_\eta^2 + \boldsymbol{a}'_{t-1}\mathrm{Cov}(\boldsymbol{\delta}_t)\boldsymbol{a}_{t-1}\\ &= \sigma_\eta^2 + (a_{t-1},\cdots,a_{t-m})\boldsymbol{\Omega}(a_{t-1},\cdots,a_{t-m})'\end{aligned} \tag{3.31}$$

记$\boldsymbol{\Omega}$的第(i,j)个元素为ω_{ij}．因为矩阵是对称的，故$\omega_{ij}=\omega_{ji}$．若$m=1$，则式(3.31)简化为$\sigma_t^2=\sigma_\eta^2+\omega_{11}a_{t-1}^2$，这是一个 ARCH(1)模型．若$m=2$，式(3.31)变为

$$\sigma_t^2 = \sigma_\eta^2 + \omega_{11}a_{t-1}^2 + 2\omega_{12}a_{t-1}a_{t-2} + \omega_{22}a_{t-2}^2$$

这不同于 ARCH(2)模型，因为有交叉乘积项$a_{t-1}a_{t-2}$．一般地，若$\boldsymbol{\Omega}$为对角阵，则 CHARMA(m)模型的条件方差等于一个 ARCH(m)模型的条件方差．因为$\boldsymbol{\Omega}$是协方差阵，它是非负定的，σ_η^2是方差、是正的，故我们有$\sigma_t^2\geqslant\sigma_\eta^2>0$对任何$t$成立．换句话说，在 CHARMA 模型下$\sigma_t^2$自动地是正数．

107

AHCH 和 CHARMA 模型的明显区别是后者在波动率方程中有a_t的延迟值的交叉乘积项．这些交叉乘积项在一些应用中会有用．例如，在资产收益率序列建模中，交叉乘积项表示前面的收益率的相互作用．可以想像，股票波动率依赖于这些相互作用．然而，交叉乘积项的个数随着阶m迅速增加，需要某些限制以使模型简单．一些可能的限制是在 CHARMA 模型中只用少数交叉乘积项．ARCH 和 CHARMA 模型的另一个区别是：高阶 CHARMA 模型的性质比 ARCH 模型的性质更难得到，因为处理随机系数要比处理常系数要难．

作为例子，我们对前面 GARCH 建模已用的标准普尔 500 指数的月超出收益率建立一个 CHARMA 模型：

$$r_t = \phi_0 + a_t, \quad a_t = \delta_{1t}a_{t-1} + \delta_{2t}a_{t-2} + \eta_t$$

所拟合的模型为

$$r_t = 0.006\,35 + a_t, \quad \sigma_t^2 = 0.001\,79 + (a_{t-1},a_{t-2})\hat{\boldsymbol{\Omega}}(a_{t-1},a_{t-2})'$$

其中

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} 0.141\,7(0.033\,3) & -0.059\,4(0.036\,5)\\ -0.059\,4(0.036\,5) & 0.308\,1(0.034\,0)\end{bmatrix},$$

括号中的数是标准差．$\hat{\boldsymbol{\Omega}}$的交叉乘积项的$t$比为$-1.63$，它是在 10%的水平下显著的．如果我们把模型改进为

$$r_t = \phi_0 + a_t, \quad a_t = \delta_{1t}a_{t-1} + \delta_{2t}a_{t-2} + \delta_{3t}a_{t-3} + \eta_t$$

但假定 δ_{3t} 与 $(\delta_{1t},\ \delta_{2t})$ 是不相关的，则我们得到拟合模型为

$$r_t = 0.006\,8 + a_t, \quad \sigma_t^2 = 0.001\,36 + (a_{t-1}, a_{t-2}, a_{t-3})\hat{\mathbf{\Omega}}(a_{t-1}, a_{t-2}, a_{t-3})'$$

其中 $\hat{\Omega}$ 的元素及其标准差(在括号中给出)为

$$\hat{\Sigma} = \begin{bmatrix} 0.121\,2(0.035\,5) & -0.062\,2(0.028\,3) & 0 \\ -0.062\,2(0.028\,3) & 0.191\,3(0.025\,4) & 0 \\ 0 & 0 & 0.298\,8(0.042\,0) \end{bmatrix},$$

所有的估计值在 5% 的水平下是显著的. 对该模型，$a_t = r_t - 0.006\,8$ 是月超出收益率与它的平均值的偏离. 所建立的 CHARMA 模型说明头两个延迟值之间存在某种相互作用. 实际上，波动率方程可近似地写成

108

$$\sigma_t^2 = 0.001\,36 + 0.12a_{t-1}^2 - 0.12a_{t-1}a_{t-2} + 0.19a_{t-2}^2 + 0.30a_{t-3}^2.$$

当 $a_{t-1}a_{t-2}$ 是负的时候，条件方差要稍大一些.

解释变量的效应

我们能容易地把 CHARMA 模型推广，使得模型中 r_t 的波动率依赖于某些解释变量. 设 $\{x_{it}\}_{i=1}^{m}$ 是 t 时刻可观测的 m 个解释变量. 考虑模型

$$r_t = \mu_t + a_t, \quad a_t = \sum_{i=1}^{m} \delta_{it} x_{i,t-1} + \eta_t \tag{3.32}$$

其中 $\boldsymbol{\delta}_t = (\delta_{1t},\ \cdots,\ \delta_{mt})'$ 和 η_t 是式(3.30)中定义的随机向量和随机变量. 那么，a_t 的条件方差为

$$\sigma_t^2 = \sigma_\eta^2 + (x_{1,t-1}, \cdots, x_{m,t-1})\mathbf{\Omega}(x_{1,t-1}, \cdots, x_{m,t-1})'$$

在实际应用中，解释变量可能会包含 a_t 的某些延迟值.

3.9 随机系数的自回归模型

文献中，随机系数自回归(RCA)模型是为了考虑不同体制间的变化而提出的，类似于计量经济学中的板条数据分析(panel data analysis)和统计学中的分等级模型(hierarchical model). 我们把 RCA 模型当作条件异方差模型来阐述，但历史上，是考虑到参数随时间演变而用它来得到过程的条件均值方程的更好描述. 如果一个时间序列 $\{r_t\}$ 满足

$$r_t = \phi_0 + \sum_{i=1}^{p} (\phi_i + \delta_{it}) r_{t-i} + a_t \tag{3.33}$$

其中 p 是正整数，$\{\boldsymbol{\delta}_t\} = \{(\delta_{1t},\ \cdots,\ \delta_{pt})'\}$ 是一列独立的均值为 0、协方差阵为 $\mathbf{\Omega}_\delta$ 的随机向量，且 $\{\boldsymbol{\delta}_t\}$ 与 $\{a_t\}$ 是独立的，则称 $\{r_t\}$ 服从 RCA(p)模型. 关于该模型更详细的讨论，请参见 Nicholls 和 Quinn(1982). 式(3.33)中的 RCA 模型的条件均值和条件方差为

109

$$\mu_t = E(r_t \mid F_{t-1}) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i}, \quad \sigma_t^2 = \sigma_a^2 + (r_{t-1}, \cdots, r_{t-p})\mathbf{\Omega}_\delta(r_{t-1}, \cdots, r_{t-p})',$$

与 CHARMA 模型形式上相同. 但 RCA 和 CHARMA 模型间存在微妙的差别：对 RCA 模型，波动率是观察到的延迟值 r_{t-i}的二次函数，而在 CHARMA 模型中，波动率是延迟的新息 a_{t-i}的二次函数.

3.10 随机波动率模型

另一种描述金融时间序列波动率演变的方法是：对 a_t 的条件方差方程引进一个新息. 参见 Melino 和 Turnbull(1990)、Harvey，Ruiz 和 Shephard(1994)、Jacquier，Polson 和 Rossi (1994). 结果得到的是随机波动率(SV)模型. 与 EGARCH 模型相似，为了保证条件方差为正，SV 模型用 $\ln(\sigma_t^2)$而不是σ_t^2. SV 模型定义为

$$a_t = \sigma_t \epsilon_t, \quad (1-\alpha_1 B - \cdots - \alpha_m B^m)\ln(\sigma_t^2) = \alpha_0 + v_t \tag{3.34}$$

其中 ϵ_t 是 iid $N(0, 1)$的，v_t 是 iid $N(0, \sigma_v^2)$的，且$\{\epsilon_t\}$和$\{v_t\}$是相互独立的，α_0 是常数，多项式 $1-\sum_{i=1}^{m}\alpha_i B^i$ 的所有根的模大于 1. 引进新息 v_t 很大程度地增加了描述σ_t^2 演变的模型的灵活性，但也增加了参数估计的困难. 为了估计 SV 模型，我们需要用到用 Kalman 滤波的伪似然(quasi-likelihood)方法或者蒙特卡罗方法. Jacquier，Polson 和 Rossi(1994)给出了伪似然和蒙特卡罗马尔可夫链(MCMC)方法的估计结果之间的比较. 估计 SV 模型比较困难是可以想像的，因为对于每个抖动此模型中用了两个新息 ϵ_t 和 v_t. 在第 10 章中将讨论一个 MCMC 方法去估计 SV 模型. 关于随机波动率模型的讨论，请参考 Taylor(1994).

Jacquier，Polson 和 Rossi(1994)的附录中给出了一些 SV 模型当 $m=1$ 的性质. 例如，当 $m=1$ 时，有

$$\ln(\sigma_t^2) \sim N\left(\frac{\alpha_0}{1-\alpha_1}, \frac{\sigma_v^2}{1-\alpha_1^2}\right) \equiv N(\mu_h, \sigma_h^2),$$

且 $E(a_t^2)=\exp(\mu_h+\sigma_h^2/2)$，$E(a_t^4)=3\exp[2\mu_h^2+2\sigma_h^2]$，和 $\mathrm{corr}(a_t^2, a_{t-i}^2)=[\exp(\sigma_h^2\alpha_1^i)-1]/[3\exp(\sigma_h^2)-1]$. 有限的经验告诉我们 SV 模型常常在模型的拟合上有改进，但在样本以外的波动率预测上与其他模型相比有时好，有时差.

3.11 长记忆随机波动率模型

最近，有人利用分数差分方法，进一步推广了 SV 模型，允许波动率有长记忆性. 正如第 2 章所讨论的，一个时间序列是一个长记忆过程，如果当间隔增加时自相关函数以幂函数(而不是以指数速度)衰减. 在波动率的研究中引进长记忆模型的动机是这样一个事实：虽然资产收益率序列本身没有前后相关性，但收益率的绝对值或平方序列的自相关函数常常是衰减很慢的. (参见 Ding，Granger 和 Engle(1993)). 图 3-9 所示的是 IBM 股票和标准普尔 500 指数从 1962 年 7 月 3 日至 1997 年 12 月 31 日的日绝对值收益率的对数序列的样本自相关函数. 这两个样本 ACF 都是正的，取值大小中等，但衰减很慢.

110

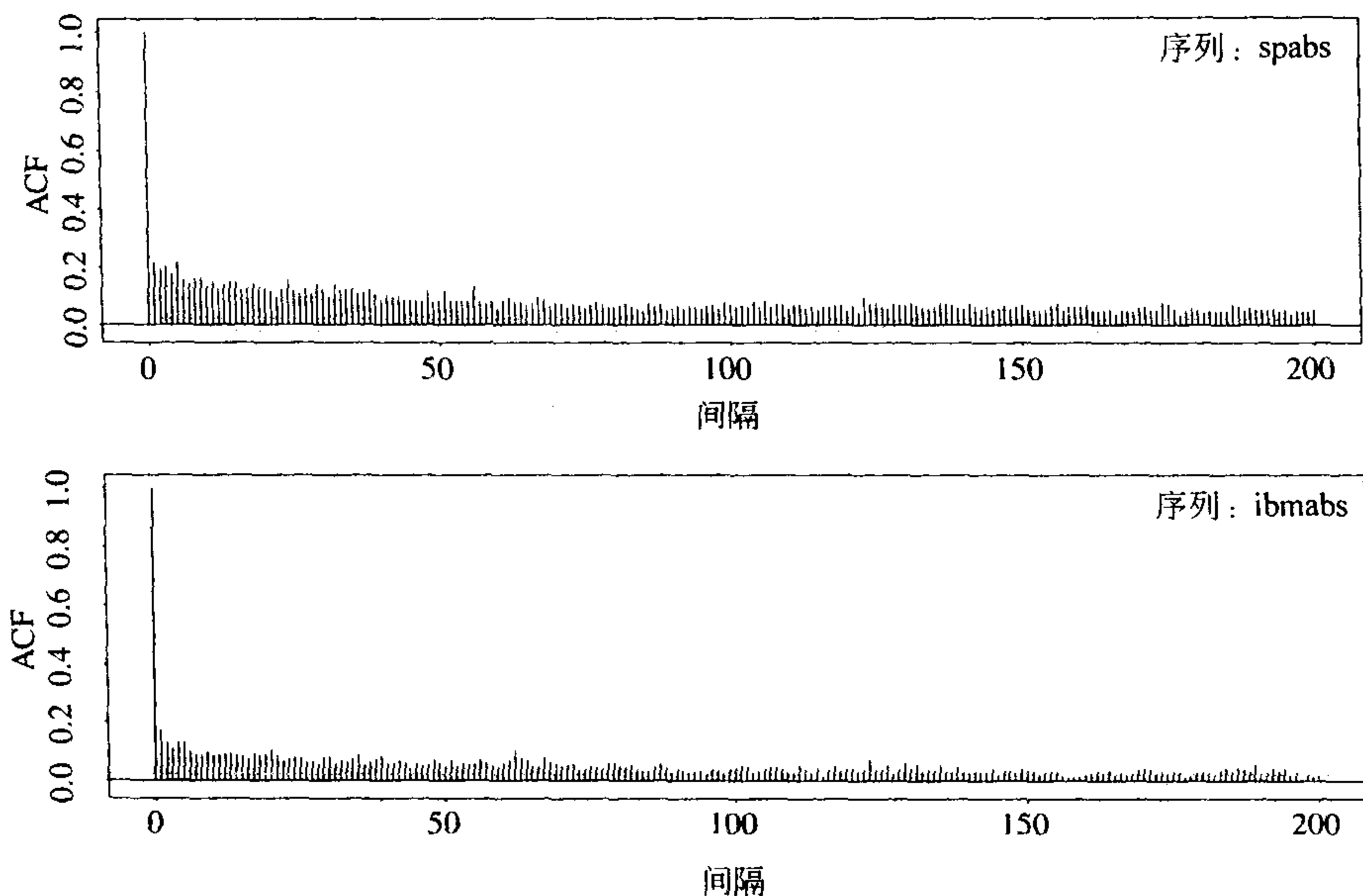


图 3-9 标准普尔 500 指数和 IBM 股票日对数收益率绝对值序列的样本 ACF，时间区间是 1962 年 7 月 3 日至 1997 年 12 月 31 日．两条水平线表示渐近的 5%上、下限

简单的长记忆随机波动率(LMSV)模型可写成

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t = \sigma \exp(u_t/2), \quad (1-B)^d u_t = \eta_t \tag{3.35}$$

其中 $\sigma>0$，ε_t 是 iid $N(0, 1)$，η_t 是 iid $N(0, \sigma_\eta^2)$且和 ε_t 相互独立，$0<d<0.5$. 长记忆特征源于分数差分$(1-B)^d$，它可推出随时间间隔的增加 u_t 的 ACF 以幂函数(而不是以指数函数)衰减．对模型(3.35)，我们有

111

$$\begin{aligned}\ln(a_t^2) &= \ln(\sigma^2) + u_t + \ln(\varepsilon_t^2) \\ &= [\ln(\sigma^2) + E(\ln\varepsilon_t^2)] + u_t + [\ln(\varepsilon_t^2) - E(\ln\varepsilon_t^2)] \\ &\equiv \mu + u_t + e_t\end{aligned}$$

这样，$\ln(a_t^2)$序列是一个高斯长记忆过程加上一个非高斯白噪声，见 Breidt，Crato 和 de Lima (1998)．长记忆随机波动率模型的估计是复杂的，但分数差分参数 d 是可以用伪最大似然法或回归方法来估计的．对标准普尔 500 指数中运用的公司股票的日收益率，取平方后再取对数，利用这样的一些序列，Bollerslev 和 Jubinski(1999)、Ray 和 Tsay(2000)发现 d 的中位数估计约为 0.38．Ray 和 Tsay(2000)以各种不同特征对公司进行分类，研究了各类公司股票的日波动率的共同的长记忆成分．他们发现同一行业公司会有更趋相同的长记忆成分(例如，美国大的国有银行和金融机构)．

3.12 另一种方法

French，Schwert 和 Stambaugh(1987)考虑另一种估计波动率的方法，他们用高频数据去计算低频收益率的波动率．近年来，这种方法已经有了一些吸引力，因为高频金融数据越来越容易获得，尤其是在外汇市场上．(见 Andersen，Bollerslev，Diebold 和 Labys(1999))．

假设我们对某资产的月波动率感兴趣，而我们可以获得该资产的日收益率．设 r_t^m 是该资产的第 t 个月的对数收益率，假设第 t 月有 n 个交易日，这个月中资产的日对数收益率为 $\{r_{t,i}\}_{i=1}^n$．利用对数收益率的性质，我们有

$$r_t^m = \sum_{i=1}^n r_{t,i}$$

假设条件方差和协方差是存在的，则

$$\mathrm{Var}(r_t^m \mid F_{t-1}) = \sum_{i=1}^n \mathrm{Var}(r_{t,i} \mid F_{t-1}) + 2\sum_{i<j}\mathrm{Cov}[(r_{t,i}, r_{t,j}) \mid F_{t-1}], \tag{3.36}$$

其中 F_{t-1} 是第 $t-1$ 月已知的信息．如果添加另外的假定，上式可以简化．例如，若假定 $\{r_{t,i}\}$ 是白噪声序列，则

$$\mathrm{Var}(r_t^m \mid F_{t-1}) = n\mathrm{Var}(r_{t,1})$$

112

其中 $\mathrm{Var}(r_{t,1})$ 可以用日收益率 $\{r_{t,i}\}_{i=1}^n$ 估计出来：

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2}{n-1}$$

其中 $\bar{r}_t$ 是第 t 月中的日对数收益率的样本均值$\left(即 \bar{r}_t = \sum_{i=1}^n r_{t,i}/n\right)$．从而，月波动率的估计为

$$\hat{\sigma}_m^2 = \frac{n}{n-1}\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2 \tag{3.37}$$

若 $\{r_{t,i}\}$ 服从一个 MA(1)模型，则

$$\mathrm{Var}(r_t^m \mid F_{t-1}) = n\mathrm{Var}(r_{t,1}) + 2(n-1)\mathrm{Cov}(r_{t,1}, r_{t,2}),$$

它的估计可由下式给出：

$$\hat{\sigma}_m^2 = \frac{n}{n-1}\sum_{i=1}^n (r_{t,i} - \bar{r}_t)^2 + 2\sum_{i=1}^{n-1}(r_{t,i} - \bar{r}_t)(r_{t,i+1} - \bar{r}_t) \tag{3.38}$$

上述波动率估计的方法是简单的，但在实际应用中会遇到一些困难．第一，对日收益率 $\{r_{t,i}\}$ 的模型是未知的，这就使式(3.36)中的协方差的估计变复杂了．第二，一个月大约 21 个交易日，是个小样本，其精确性值得怀疑．估计的精确性取决于 $\{r_{t,i}\}$ 的动态结构和它的分布，若日对数收益率有较高的超出峰度和较强的前后相关性，则式(3.37)和式(3.38)中的样本估计 $\hat{\sigma}_m^2$ 甚至是不相合的(见 Bai，Russell 和 Tiao(2000))．为了使这种方法更有实用价值，还需要更进一步的研究．

例 3.4 考虑标准普尔 500 指数的对数收益率的月波动率，时间是从 1980 年 1 月至 1999 年 12 月. 我们用三种方法来计算波动率. 第一种方法，用日对数收益率和式(3.37)(假定日对数收益率是白噪声序列). 第二种方法，用日收益率但假定一个 MA(1)模型(即用式(3.38)). 第三种方法是对 1962 年 1 月至 1999 年 12 月的月收益率用一个 GARCH(1, 1)模型. 我们使用更长的时间段是为了得到更精确的月波动率估计. 所用的 GARCH(1, 1)模型为

$$r_t^m = 0.658 + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 3.349 + 0.086 a_{t-1}^2 + 0.735 \sigma_{t-1}^2$$

其中 ε_t 是标准高斯白噪声序列. 图 3-10 所示的是估计出的月波动率的时间图，我们清楚地看到基于日收益率的波动率估计值要比基于月收益率和 GARCH(1, 1)模型的波动率估计值大许多. 特别地，当使用日收益率时，1987 年 10 月的波动率估计值大约为 680. 而为了三个图有相同的尺度范围，我们对它们进行了截断.

113

注释: 在式(3.37)中，如果我们进一步假定样本均值 $\bar{r}_t$ 是 0，则 $\hat{\sigma}_m^2 \approx \sum_{i=1}^{n} r_{t,i}^2$. 这时，一个月中日对数收益率的平方和就可以作为月波动率的估计.

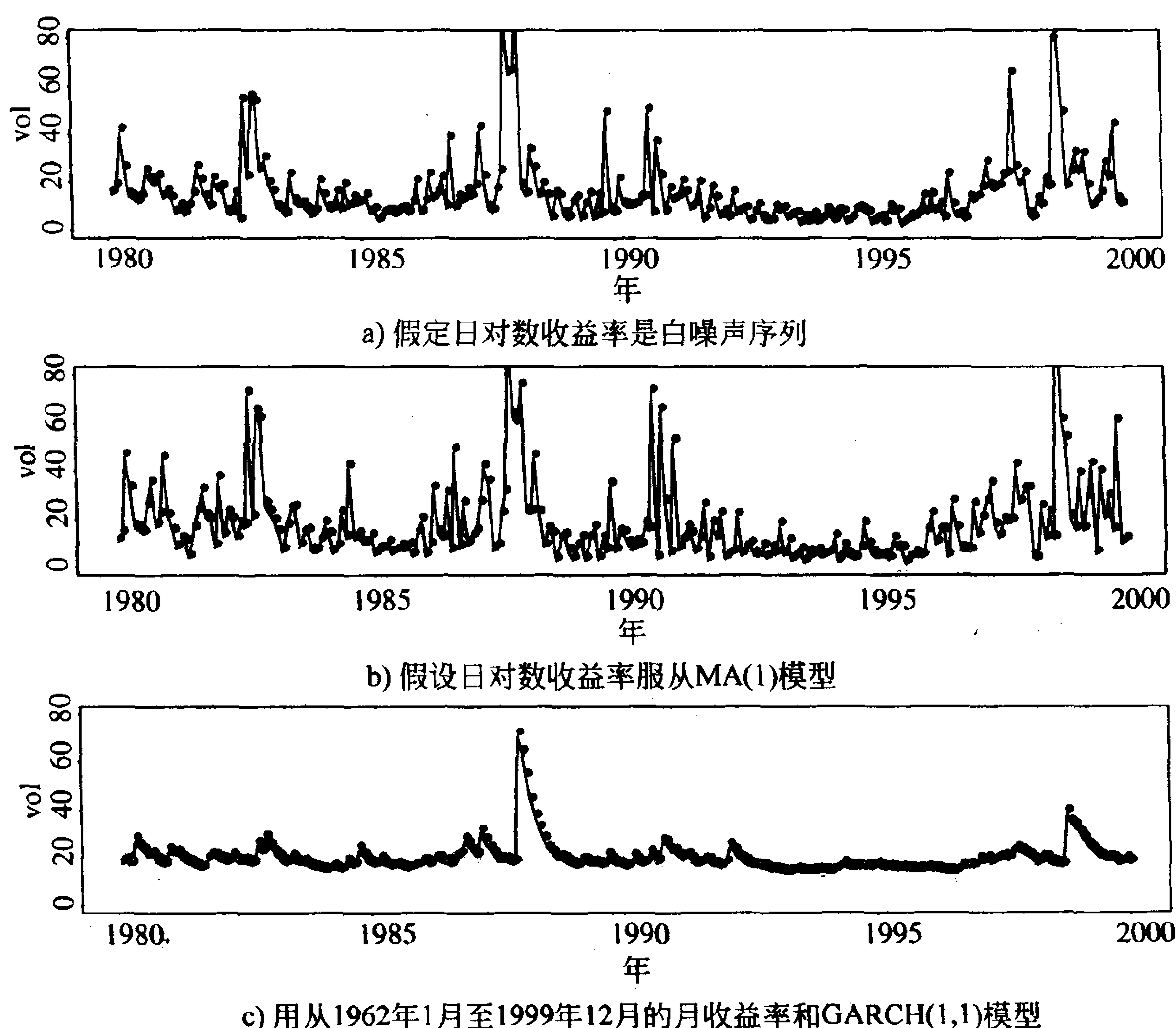


图 3-10 标准普尔 500 指数对数收益率的月波动率估计值的时间图，时间范围是 1980 年 1 月至 1999 年 12 月

3.13 应用

在这一节中，我们运用这一章中讨论过的波动率模型来研究有实际重要性的一些问题. 所用数据是从 1926 年 1 月至 1999 年 12 月 IBM 股票和标准普尔 500 指数的月对数收益率，共有 888 个观察值，收益率是以百分比形式给出并包括了分红. 图 3-11 所示的是这两个收益率序列的时间图.

114

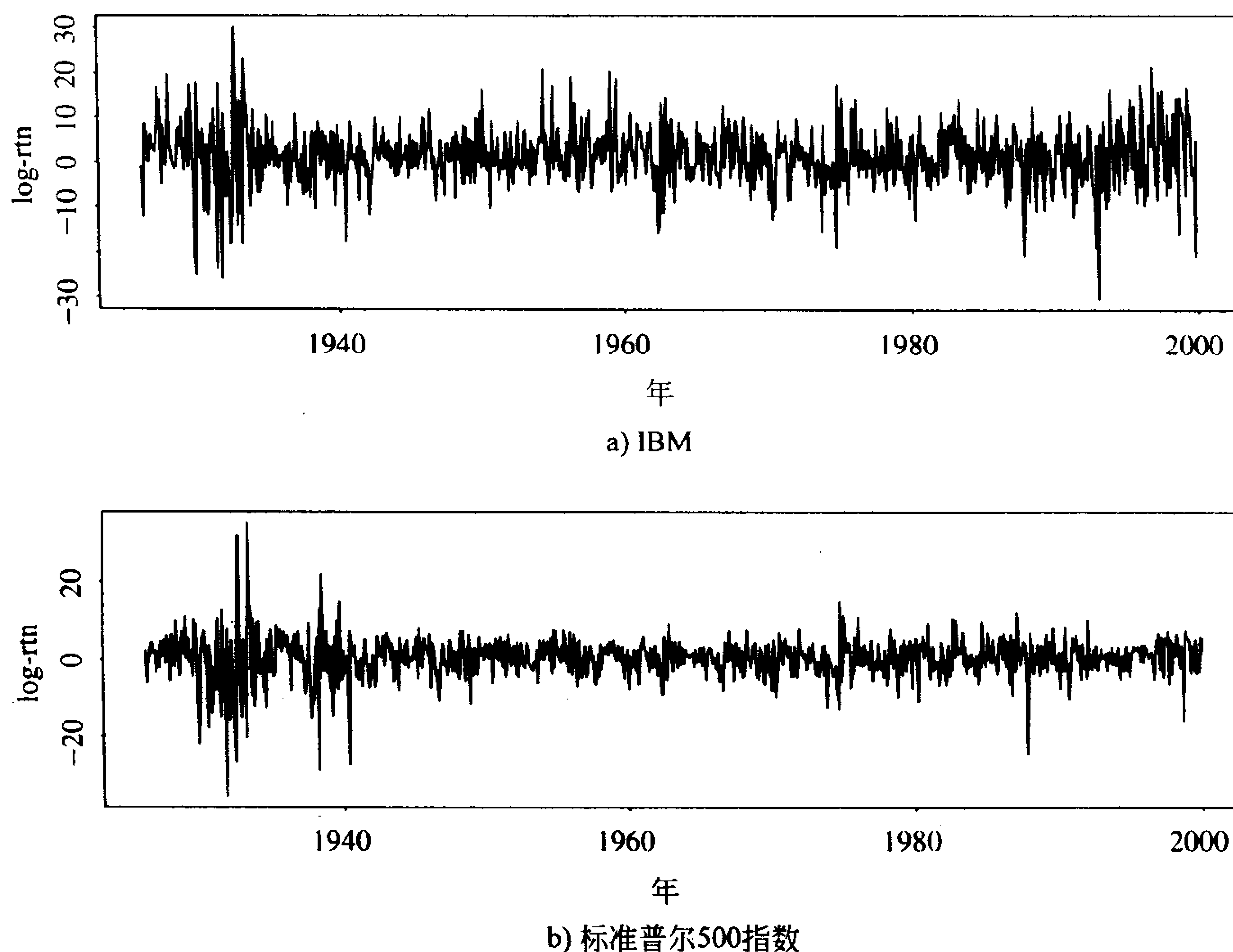


图 3-11 IBM 股票和标准普尔 500 指数月对数收益率的时间图. 样本时间段是从 1926 年 1 月至 1999 年 12 月. 收益率是以百分比给出，并包括了分红

例 3.5 我们关心这样的问题：是否一支股票的日波动率在夏季比较低？如果是，低多少？对这两个问题肯定的回答在股票期权定价中有实际意义. 我们用图 3-11a 所示的 IBM 股票的月对数的收益率作为例子来说明怎样回答这两个问题.

记月收益率序列为 r_t. 给数据拟合一个高斯 GARCH(1, 1)模型：

$$r_t = 1.23 + 0.099 r_{t-1} + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 3.206 + 0.103 a_{t-1}^2 + 0.825 \sigma_{t-1}^2 \tag{3.39}$$

均值方程中的两个参数估计值的标准差分别为 0.222 和 0.037，波动率方程中三个参数估计值的标准差分别为 0.947，0.021，0.037. 对标准化的残差 $\tilde{a}_t = a_t/\sigma_t$，Ljung-Box 统计量的值 $Q(10)=7.82(0.553)$ 和 $Q(20)=21.22(0.325)$，括号中为 p 值. 这样，均值方程的残差没有

115 前后相关性. 而对$\tilde{a}_t^2$，$Q(10)=2.89(0.98)$和$Q(20)=7.26(0.99)$，表明标准化残差中无条件异方差性. 所拟合的模型是合适的，这个模型是进一步研究的出发点.

为了研究资产的股票波动率的夏季效益，先定义一个示性变量：

$$u_t = \begin{cases} 1 & \text{若 } t \text{ 表示六月、七月、八月} \\ 0 & \text{其他} \end{cases} \tag{3.40}$$

并且把波动率方程修改为

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + u_t(\alpha_{00} + \alpha_{10} a_{t-1}^2 + \beta_{10} \sigma_{t-1}^2)$$

这个方程用两个 GARCH(1，1)模型来描述一个股票收益率的波动率：一个模型针对夏季的月份，另一个针对剩下的月份. 对 IBM 股票的月对数收益率，α_{10}和β_{10}的估计值在 10%的水平下是不显著的，因此我们对方程进行改进，得到模型

$$r_t = 1.21 + 0.099 r_{t-1} + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 4.539 + 0.113 a_{t-1}^2 + 0.816 \sigma_{t-1}^2 - 5.154 u_t \tag{3.41}$$

均值方程中参数估计值的标准差分别是 0.218 和 0.037，波动率方程中参数估计的标准差分别是 1.071、0.022、0.037 和 1.900. 标准化残差$\tilde{a}_t = a_t/\sigma_t$ 的 Ljung-Box 统计量的值 $Q(10)=7.66(0.569)$和 $Q(20)=21.64(0.302)$，因此，在标准化残差序列中没有前后相关性. 而对$\tilde{a}_t^2$的 Ljung-Box 统计量 $Q(10)=3.38(0.97)$、$Q(20)=6.82(0.99)$，显示出在标准化残差中也没有条件异方差性. 改进后的模型是合适的.

比较式(3.39)和式(3.41)中的波动率模型，我们会得出如下结论：第一，因为系数-5.514的 p 值为 0.006 7，在 1%的水平下是统计显著的，因此股票波动率的夏季效应是显著的. 另外，估计值的负号证实了 IBM 股票月对数收益率的波动率在夏季确实较低. 第二，把式(3.41)中波动率模型改写成

$$\sigma_t^2 = \begin{cases} -0.615 + 0.113 a_{t-1}^2 + 0.816 \sigma_{t-1}^2 & \text{若 } t \text{ 是六月、七月、八月份} \\ 4.539 + 0.113 a_{t-1}^2 + 0.816 \sigma_{t-1}^2 & \text{其他} \end{cases}$$

负常数项$-0.615=4.539-5.514$ 与直观不一致. 然而，因为 4.539 和 5.514 的标准差相对较大，估计值之差-0.615 会不是显著地不同于 0. 为了证实这一点，我们限制波动率方程中的

116 常数项对夏季月份是 0，然后重新拟合模型. 利用如下方程

$$\sigma_t^2 = \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma(1-u_t)$$

所拟合模型为

$$r_t = 1.21 + 0.099 r_{t-1} + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 0.114 a_{t-1}^2 + 0.811 \sigma_{t-1}^2 + 4.552(1-u_t) \tag{3.42}$$

均值方程中参数估计的标准差分别为 0.219 和 0.038，波动率方程中参数估计的标准差分别为 0.022、0.034 和 1.094. 标准化残差$\tilde{a}_t$ 的 Ljung-Box 统计量 $Q(10)=7.68$ 和 $Q(20)=21.67$，$\tilde{a}_t^2$ 序列的 Ljung-Box 统计量 $Q(10)=3.17$ 和 $Q(20)=6.85$. 检验统计量的这些值与前面得到的值很靠近，都是在 5%水平下不显著的.

式(3.42)的波动率可用来评价夏季效应在 IBM 股票波动率上的体现. 基于式(3.42)，对

1999 年的 IBM 月对数收益率来说，a_t^2 和 σ_t^2 的中位数分别是 29.4 和 75.1. 利用这些值，我们有：对夏季 $\sigma_t^2=0.114\times29.4+0.811\times75.1=64.3$，对其他月份 $\sigma_t^2=68.8$. 这两个波动率之比为 64.3/68.8≈93%. 也就是说，在夏季 IBM 股票的月对数收益率的波动率减小 7%。

例 3.6 标准普尔 500 指数在衍生产品市场上被广泛应用，因此对它的波动率建模就是一个研究热点. 此例中，我们所问的问题是：组成标准普尔 500 指数的单个股票过去的收益率对该指数现在的波动率建模是否有影响？这个问题的详细研究超出了本章的范围，这里我们仅用 IBM 股票过去的收益率作为解释变量来说明一下问题.

数据如图 3-11 所示，标准普尔 500 指数的月收益率序列记为 r_t. 利用 r_t 序列和高斯 GARCH 模型，我们得到如下特殊的 GARCH(2, 1)模型

$$r_t = 0.609 + a_t,\quad a_t = \sigma_t\epsilon_t,\quad \sigma_t^2 = 0.717 + 0.147a_{t-2}^2 + 0.839\sigma_{t-1}^2 \tag{3.43}$$

均值方程中常数项的标准差为 0.138，波动率方程中参数的标准差分别为 0.214、0.021 和 0.017. 基于标准化残差$\tilde{a}_t=a_t/\sigma_t$，$Q(10)=11.51(0.32)$，$Q(20)=23.71(0.26)$，括号内为 p 值. 对$\tilde{a}_t^2$ 序列，$Q(10)=9.42(0.49)$，$Q(20)=13.01(0.88)$. 从而，在 5%的水平下，模型是合适的.

下面我们评价 IBM 股票过去的收益率对标准普尔 500 指数波动率建模的影响. 我们把波动率方程修改为

$$\sigma_t^2 = \alpha_0 + \alpha_2a_{t-2}^2 + \beta_1\sigma_{t-1}^2 + \gamma(x_{t-1} - 1.24)^2$$

117

其中 x_t 是 IBM 股票的月对数收益率，1.24 是 x_t 的样本均值. 对 r_t 拟合的模型变为

$$\begin{aligned} r_t &= 0.616 + a_t,\quad a_t = \sigma_t\epsilon_t \\ \sigma_t^2 &= 1.069 + 0.148a_{t-2}^2 + 0.834\sigma_{t-1}^2 - 0.007(x_{t-1} - 1.24)^2 \end{aligned} \tag{3.44}$$

均值方程中参数估计的标准差为 0.139，波动率方程中各参数估计的标准差分别为 0.271、0.020、0.018 和 0.002. 关于模型的检验，对标准化残差$\tilde{a}_t=a_t/\sigma_t$，$Q(10)=11.39(0.33)$，$Q(20)=23.63(0.26)$，而对$\tilde{a}_t^2$ 序列，$Q(10)=9.35(0.50)$，$Q(20)=13.51(0.85)$. 因此，此模型是合适的.

因为检验 $\gamma=0$ 的 p 值为 0.003 9，所以延迟间隔为 1 的 IBM 股票收益率对标准普尔 500 指数波动率的影响在 1%的水平下是显著的. 系数的符号是负的，这可推出延迟间隔为 1 的 IBM 股票过去的收益率能减小标准普尔 500 指数波动率，这一点是可以理解的. 表 3-3 给出了利用式(3.43)和式(3.44)的模型给标准普尔 500 指数拟合的从 1999 年 7 月至 12 月波动率. 从表中可看出，IBM 股票对数收益率的过去值对标准普尔 500 指数波动率确实有影响.

表 3-3

月	7/99	8/99	9/99	10/99	11/99	12/99
模型(3.43)	26.30	26.01	24.73	21.69	20.71	22.46
模型(3.44)	23.32	23.13	22.46	20.00	19.45	18.27

注：用带和不带 IBM 股票过去对数收益率的模型拟合的标准普尔 500 指数的月对数收益率从 1999 年 7 月至 12 月的波动率.

3.14 GARCH 模型的峰度

波动率估计中的不确定性是一个重要的论题，但它经常被忽视. 为了评估被估波动率的易变性，人们必须考虑波动率模型的峰度. 在这一节中，我们导出 GARCH(1，1)模型的超出峰度，同样的思想可应用到其他的 GARCH 模型. 所考虑的模型为

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

其中 $\alpha_0>0$，$\alpha_1\geqslant 0$，$\beta_1\geqslant 0$，$\alpha_1+\beta_1<1$，$\{\varepsilon_t\}$是 iid 序列，满足

118

$$E(\varepsilon_t) = 0, \quad \mathrm{Var}(\varepsilon_t) = 1, \quad E(\varepsilon_t^4) = K_\varepsilon + 3,$$

其中 K_ε 是新息 ε_t 的超出峰度. 基于以上假设，我们有

- $\mathrm{Var}(a_t)=E(\sigma_t^2)=\alpha_0/[1-(\alpha_1+\beta_1)]$.
- $E(a_t^4)=(K_\varepsilon+3)E(\sigma_t^4)$，只要 $E(\sigma_t^4)$存在.

在波动率的模型两边取平方，有

$$\sigma_t^4 = \alpha_0^2 + \alpha_1^2 a_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2\alpha_0\alpha_1 a_{t-1}^2 + 2\alpha_0\beta_1\sigma_{t-1}^2 + 2\alpha_1\beta_1\sigma_{t-1}^2 a_{t-1}^2.$$

再两边取期望并利用前面所述的性质，我们有

$$E(\sigma_1^4) = \frac{\alpha_0^2(1+\alpha_1+\beta_1)}{[1-(\alpha_1+\beta_1)][1-\alpha_1^2(K_\varepsilon+2)-(\alpha_1+\beta_1)^2]}$$

只要 $1>\alpha_1+\beta_1>0$，且 $1-\alpha_1^2(K_\varepsilon+2)-(\alpha_1+\beta_1)^2>0$. 若 a_t 的超出峰度存在，则它是

$$\begin{aligned} K_a &= \frac{E(a_t^4)}{[E(a_t^2)]^2} - 3 \\ &= \frac{(K_\varepsilon+3)[1-(\alpha_1+\beta_1)^2]}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2-K_\varepsilon\alpha_1^2} - 3 \end{aligned}$$

此超出峰度能写成一个非常有启发性的式子. 首先，考虑 ε_t 是正态分布的情形. 在这个情形下，$K_\varepsilon=0$，并且通过简单的代数运算可得

$$K_a^{(g)} = \frac{6\alpha_1^2}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2},$$

其中上角标(g)用来表示高斯分布. 这个结果有两个重要推论：(a)若 $1-2\alpha_1^2-(\alpha_1+\beta_1)^2>0$，则 a_t 的峰度存在；(b)若 $\alpha_1=0$，则 $K_a^{(g)}=0$，意味着对应的 GARCH(1，1)模型没有厚尾.

其次，考虑 ε_t 不是高斯的情形. 利用前面的结果，我们有

119

$$\begin{aligned} K_a &= \frac{K_\varepsilon - K_\varepsilon(\alpha_1+\beta_1) + 6\alpha_1^2 + 3K_\varepsilon\alpha_1^2}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2-K_\varepsilon\alpha_1^2} \\ &= \frac{K_\varepsilon[1-2\alpha_1^2-(\alpha_1+\beta_1)^2] + 6\alpha_1^2 + 5K_\varepsilon\alpha_1^2}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2-K_\varepsilon\alpha_1^2} \\ &= \frac{K_\varepsilon + K_a^{(g)} + \frac{5}{6}K_\varepsilon K_a^{(g)}}{1-\frac{1}{6}K_\varepsilon K_a^{(g)}} \end{aligned}$$

这个结果最初由 George C. Tiao 得到(见 Bai，Russell 和 Tiao(2001))，它对所有存在峰度的

GARCH 模型成立. 例如，当 $\beta_1=0$ 时，模型化为 ARCH(1)模型，这时 $K_a^{(g)}=6\alpha_1^2/(1-3\alpha_1^2)$，只要 $1>3\alpha_1^2$，并且 a_t 的超出峰度为

$$K_a=\frac{(K_\varepsilon+3)(1-\alpha_1^2)}{1-(K_\varepsilon+3)\alpha_1^2}-3=\frac{K_\varepsilon+2K_\varepsilon\alpha_1^2+6\alpha_1^2}{1-3\alpha_1^2-K_\varepsilon\alpha_1^2}$$
$$=\frac{K_\varepsilon(1-3\alpha_1^2)+6\alpha_1^2+5K_\varepsilon\alpha_1^2}{1-3\alpha_1^2-K_\varepsilon\alpha_1^2}$$
$$=\frac{K_\varepsilon+K_a^{(g)}+\frac{5}{6}K_\varepsilon K_a^{(g)}}{1-\frac{1}{6}K_\varepsilon K_a^{(g)}}$$

上述结果表明：对 GARCH(1，1)模型来说，系数 α_1 在决定 a_t 的尾部行为时起关键作用. 若 $\alpha_1=0$，则 $K_a^{(g)}=0$ 且 $K_a=K_\varepsilon$，这时 a_t 的尾部行为与标准化噪声 ε_t 的尾部行为相似. 若 $\alpha_1>0$，则 $K_a^{(g)}>0$，a_t 过程有厚尾性.

对一个（标准化的）自由度为 v 的学生-t 分布，若 $v>4$，我们有 $E(\varepsilon_t^4)=6/(v-4)+3$. 因此，对 $v>4$，ε_t 的超出峰度为 $K_\varepsilon=6/(v-4)$. 这就是在这一章中当 t 分布的自由度是事先给定时，我们使用 t_5 的部分原因. a_t 的超出峰度为 $K_a=[6+(v+1)K_a^{(g)}]/[v-4-K_a^{(g)}]$，只要 $1-2\alpha_1^2(v-1)/(v-4)-(\alpha_1+\beta_1)^2>0$.

附录 A 估计波动率模型的一些 RATS 程序

在说明过程中所用的数据文件是“sp500.dat”，该文件中包含的是标准普尔 500 指数的月超出收益率，共 792 个观察值. RATS 程序中的注解前面带 * 号.

A. 带常数均值方程的 GARCH(1，1)模型

```
all 0  792:1
open data sp500.dat
data(org=obs) / rt
*** initialize the conditional variance function
set h = 0.0
*** specify the parameters of the model
nonlin mu a0 a1 b1
*** specify the mean equation
frml at = rt(t)-mu
*** specify the volatility equation
frml gvar = a0+a1*at(t-1)**2+b1*h(t-1)
*** specify the log likelihood function
frml garchln = -0.5*log(h(t)=gvar(t))-0.5*at(t)**2/h(t)
*** sample period used in estimation
smpl 2 792
*** initial estimates
compute a0 = 0.01, a1 = 0.1, b1 = 0.5, mu = 0.1
maximize(method=bhhh,recursive,iterations=150) garchln
set fv = gvar(t)
set resid = at(t)/sqrt(fv(t))
```

120

```
set residsq = resid(t)*resid(t)
*** Checking standardized residuals
cor(qstats,number=20,span=10) resid
*** Checking squared standardized residuals
cor(qstats,number=20,span=10) residsq
```

B. 带学生-t 分布的新息的 GARCH(1，1)模型

```
all 0  792:1
open data sp500.dat
data(org=obs) / rt
set h = 0.0
nonlin mu a0 a1 b1 v
frml at = rt(t)-mu
frml gvar = a0+a1*at(t-1)**2+b1*h(t-1)
frml tt = at(t)**2/(h(t)=gvar(t))
frml tln = %LNGAMMA((v+1)/2.)-%LNGAMMA(v/2.)-0.5*log(v-2.)
frml gln = tln-((v+1)/2.)*log(1.0+tt(t)/(v-2.0))-0.5*log(h(t))
smpl  2 792
compute a0 = 0.01, a1 = 0.1, b1 = 0.5, mu = 0.1, v = 10
maximize(method=bhhh,recursive,iterations=150) gln
set fv = gvar(t)
set resid = at(t)/sqrt(fv(t))
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

C. 对 IBM 股票月对数收益率的 AR(1)-EGARCH(1，0)模型

```
all 0  864:1
open data m-ibm.dat
data(org=obs) / rt
set h = 0.0
nonlin c0 p1 th ga a0 a1
frml at = rt(t)-c0-p1*rt(t-1)
frml epsi = at(t)/(sqrt(exp(h(t))))
frml g = th*epsi(t)+ga*(abs(epsi(t))-sqrt(2./%PI))
frml gvar = a1*h(t-1)+(1-a1)*a0+g(t-1)
frml garchln = -0.5*(h(t)=gvar(t))-0.5*epsi(t)**2
smpl 3  864
compute c0 = 0.01, p1 = 0.01, th = 0.1, ga = 0.1
compute a0 = 0.01, a1 = 0.5
maximize(method=bhhh,recursive,iterations=150) garchln
set fv = gvar(t)
set resid = epsi(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

[121]

练习题

1. 对 GARCH(1，2)模型，导出以 h 为预测原点的向前多步预测公式.
2. 对 GARCH(2，1)模型，导出以 h 为预测原点的向前多步预测公式.

3. 假定 r_1，…，r_n 是来自于服从如下 AR(1)-GARCH(1，1)模型的收益率序列的观察值，$r_t=\mu+\phi_1 r_{t-1}+a_t$，$a_t=\sigma_t\epsilon_t$，$\sigma_t^2=\alpha_0+\alpha_1 a_{t-1}^2+\beta_1\sigma_{t-1}^2$，其中 ϵ_t 是标准的高斯白噪声序列．导出这组数据的条件对数似然函数．
4. 在上题中，假定 ϵ_t 服从标准化的自由度为 v 的学生-t 分布．导出数据的条件对数似然函数．
5. 考虑 Intel 股票从 1973 年至 1997 年的月对数收益率．对该序列建立一个 GARCH 模型并计算以 1997 年 12 月为预测原点的向前 1 步至 5 步的波动率预测值．
6. 文件"m-mrk. dat"包含的 Merck 股票的月简单收益率，有三列：月简单收益率、年度、月份．把简单收益率变换成对数收益率．
   - 此对数收益率存在 ARCH 效应吗？用 Ljung-Box 统计量，对收益率平方的 5 个间隔和 10 个间隔的自相关系数，在 5%的显著水平下回答．
   - 用对数收益率的平方的 PACF(样本偏自相关函数)识别一个 ARCH 模型，然后给数据拟合被识别的模型，写出所拟合的模型．
7. 文件"m-mmm. dat"包括两列，分别是 3M 公司股票的月简单收益率和日期．把收益率变换成对数收益率．

   122

   - 对数收益率存在 ARCH 效应吗？用 5 个间隔和 10 个间隔的自相关系数所构成的 Ljung-Box 统计量在 5%的显著水平下来回答．
   - 用收益率平方的 PACF 识别一个 ARCH 模型．所拟合的模型是什么？
   - 共有 623 个数据点．利用所拟合的模型来预测 $t=624$ 和 $t=625$ 时的波动率(预测原点为 $h=623$)．
   - 对 3M 股票的对数收益率建立一个 ARCH-M 模型．在 5%的显著水平下检验风险溢价为 0 的假设，得出你的结论．
   - 对 3M 股票的对数收益率建立一个 EGARCH 模型．利用所建模型计算以 $h=623$，为预测原点的向前 1 步和向前 2 步的波动率预测．
8. 文件"m-gmsp5099. dat"以百分比的形式给出 General Motors 股票和标准普尔 500 指数从 1950 年至 1999 年的月对数收益率．GM 股票收益率在第一列．
   - 给 GM 股票对数收益率建立一个带高斯新息的 GARCH 模型．检验模型并写出最后所拟合的模型．
   - 给 GM 股票的对数收益率建立一个带高斯新息的 GARCH-M 模型．所拟合的模型是什么？
   - 给 GM 股票的对数收益率建立一个带自由度为 6 的学生-t 分布的新息的 GARCH 模型，检验模型并写出最后拟合的模型．
   - 给 GM 股票的对数收益率建立一个带学生-t 分布的新息的 GARCH 模型，包括 t 分布自由度的估计．写出所拟合的模型．设 v 是学生-t 分布的自由度，在 5%的显著水平下检验假设 H_0：$v=6$ 对 H_a：$v\neq6$．
   - 给 GM 股票的对数收益率建立一个 EGARCH 模型．所拟合的模型是什么？

- 比较对 GM 股票的对数收益率所得到的所有波动率模型. 存在显著差别吗? 为什么?

9. 再考虑文件“m-gmsp5099. dat”.
   - 对标准普尔 500 指数的月对数收益率建立一个高斯 GARCH 模型. 仔细检验模型.
   - 在该指数收益率的波动率中存在夏季效应吗? 利用上小题中所建立的 GARCH 模型来回答.
   - GM 股票的延迟收益率在该指数波动率建模中有用吗? 利用第 1 小题中所建的 GARCH 模型作为基础模型来比较.

123

10. 文件“d-ibmln. dat”中含有 IBM 股票从 1962 年 7 月至 1997 年 12 月的日对数收益率，以百分数形式给出，共 8938 个观察值. 该文件只有一列. 对此序列拟合一个 GARCH(1, 1)模型. 所拟合的模型是什么?

参考文献

Andersen, T. G., and Bollerslev, T. (1998), "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, 39, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (1999), "The distribution of exchange rate volatility," Working paper, Economics Department, University of Pennsylvania.

Bai, X., Russell, J. R., and Tiao, G. C. (2000), "Beyond Metron's Utopia: effects of dependence and non-normality on variance estimates using high-frequency data," Working paper, Graduate School of Business, University of Chicago.

Bai, X., Russell, J. R., and Tiao, G. C. (2001), "Kurtosis of GARCH and stochastic volatility models," Working paper, Graduate School of Business, University of Chicago.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T. (1990), "Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach," *Review of Economics and Statistics*, 72, 498–505.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), "ARCH modeling in finance," *Journal of Econometrics*, 52, 5–59.

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994), "ARCH model" in Handbook of Econometrics IV, 2959–3038, ed. Engle, R. F., and McFadden, D. C. Amsterdam: Elsrier Science.

Bollerslev, T., and Jubinski, D. (1999), "Equality trading volume and volatility: Latent information arrivals and common long-run dependencies," *Journal of Business & Economic Statistics*, 17, 9–21.

Breidt, F. J., Crato, N., and de Lima, P. (1998), "On the detection and estimation of long memory in stochastic volatility," *Journal of Econometrics*, 83, 325–348.

Cao, C., and Tsay, R. S. (1992), "Nonlinear time series analysis of stock volatilities," *Journal of Applied Econometrics*, 7, s165–s185.

Ding, Z., Granger, C. W. J., and Engle, R, F. (1993), "A long memory property of stock returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

Engle, R. F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations," *Econometrica*, 50, 987–1007.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987), "Expected stock returns and volatility," *Journal of Financial Economics*, 19, 3–29.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994), "Multivariate stochastic variance models," *Review of Economic Studies*, 61, 247–264.

Jacquier, E., Polson, N. G., and Rossi, P. (1994), "Bayesian analysis of stochastic volatility models" (with discussion), *Journal of Business & Economic Statistics*, 12, 371–417.

McLeod, A. I., and Li, W. K. (1983), "Diagnostic checking ARMA time series models using squared-residual autocorrelations," *Journal of Time Series Analysis*, 4, 269–273. [124]

Melino, A., and Turnbull, S. M. (1990), "Pricing foreign currency options with stochastic volatility," *Journal of Econometrics*, 45, 239–265.

Nelson, D. B. (1990), "Stationarity and persistence in the GARCH(1, 1) model," *Econometric Theory*, 6, 318–334.

Nelson, D. B. (1991), "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59, 347–370.

Nicholls, D. F., and Quinn, B. G. (1982), *Random Coefficient Autoregressive Models: An Introduction*, Lecture Notes in Statistics, 11. Springer-Verlag: New York.

Ray, B. K., and Tsay, R. S. (2000), "Long-range dependence in daily stock volatilities," *Journal of Business & Economic Statistics*, 18, 254–262.

Taylor, S. J. (1994), "Modeling stochastic volatility," *Mathematical Finance*, 4, 183–204.

Tong, H. (1978), "On a threshold model," in *Pattern Recognition and Signal Processing*, ed. C.H. Chen, Sijhoff & Noordhoff: Amsterdam.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press: Oxford.

Tsay, R. S. (1987), "Conditional heteroscedastic time series models," *Journal of the American Statistical Association*, 82, 590–604. [125]

第4章　非线性模型及其应用

本章的论题是金融数据中的非线性问题和金融时间序列分析中有用的非线性经济计量模型. 考虑一元时间序列 x_t，为简单起见，假定观察的时间间隔是等距的. 记 $\{x_t \mid t=1, \cdots, T\}$ 为观察值，T 是样本容量. 如第 2 章中所述，一个纯随机的时间序列 x_t 称为线性的，如果它能表示成

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \tag{4.1}$$

其中 μ 是常数，$\psi_0=1$，ψ_i 是实数，$\{a_t\}$是独立同分布随机变量序列，其分布函数是合理定义的. 假定 a_t 的分布是连续的且 $E(a_t)=0$. 在许多场合下，我们进一步假定 $\mathrm{Var}(a_t)=\sigma_a^2$，甚至更强，$a_t$ 是高斯的. 若 $\sigma_a^2 \sum_{i=1}^{\infty}\psi_i^2 < \infty$，则 x_t 是弱平稳的(即 x_t 的头两阶矩是随时间不变的). 第 2 章中 ARMA 过程是线性的，因为它有形如式(4.1)的 MA 表示. 任何不能表示成式(4.1)的随机过程称为是非线性的. 这个非线性的定义是针对纯随机的时间序列. 人们可以拓展这个定义，允许 x_t的均值是一些外生变量的线性函数，这些外生变量包括时间指标和某些周期函数. 但这样的均值函数很容易用第 2 章中介绍的方法来处理，我们在这里就不讨论它了. 数学上讲，x_t 的一个纯随机时间序列模型是由现在和过去的抖动所构成的独立同分布序列的一个函数：

$$x_t = f(a_t, a_{t-1}, \cdots) \tag{4.2}$$

当 $f(.)$是其自变量的线性函数时，就是式(4.1)中的线性模型. $f(.)$中的任何非线性性都会导致一个非线性模型. 一般的非线性模型(4.2)是不能直接应用的，因为它包含了太多的参数.

为了把文献中可见到的非线性模型放在一个合适的框架中，我们把 x_t 的模型写成它的条件矩的形式. 设 F_{t-1}是由 $t-1$ 时刻已有信息产生的 σ 域，典型的 F_{t-1}是由$\{x_{t-1}, x_{t-2}, \cdots\}$和$\{a_{t-1}, a_{t-2}, \cdots\}$中的元素线性组合组成的. 给定 F_{t-1}，x_t 的条件均值和条件方差分别是

$$\mu_t = E(x_t \mid F_{t-1}) \equiv g(F_{t-1}), \quad \sigma_t^2 = \mathrm{Var}(x_t \mid F_{t-1}) \equiv h(F_{t-1}), \tag{4.3}$$

其中 $g(.)$和 $h(.)$是有意义的函数，$h(.)>0$. 这样，我们把模型限制于

$$x_t = g(F_{t-1}) + \sqrt{h(F_{t-1})}\epsilon_t$$

其中 $\epsilon_t=a_t/\sigma_t$ 是标准化的抖动. 对式(4.1)中的线性序列 x_t，$g(.)$是 F_{t-1}中元素的线性函数，$h(.)=\sigma_t^2$. 非线性模型的发展就在于式(4.3)中的两个方程的扩展. 若 $g(.)$是非线性的，x_t 称为是均值非线性的. 若 $h(.)$是随时间变化的，则 x_t 是方差非线性的. 第 3 章的条件异方差模型都是方差非线性的，因为它们的条件方差 σ_t^2 随时间变化. 事实上，除了 GARCH-M 模型中 μ_t 依赖于 σ_t^2，从而随时间变化外，第 3 章中所有的波动率模型都是着眼于式(4.3)中条件方差方程的扩展或修改. 基于熟知的 Wold 分解，一个弱平稳的、纯随机的时间序列能表示成不相

关的抖动的一个线性函数. 对于平稳波动率序列，抖动是互不相关的，但不是独立的. 本章讨论的模型是另一种向非线性的扩展，这种非线性性是由对式(4.3)中的条件均值方程的改动或扩展而引起的.

统计学文献中已有许多非线性时间序列模型被提出，如 Granger 和 Andersen (1978)的双线性模型，Tong (1978)的门限自回归(TAR)模型，Priestley (1980)的状态相依模型和 Hamilton (1989)的马尔可夫转移模型. 这些非线性模型的基本思想都是让条件均值函数 μ_t 按简单的参数的非线性函数随时间演变. 近来，利用计算设备和计算方法上的进展，出现了一些新的非线性模型. 这些扩展包括 Carlin，Polson 和 Stoffer(1992) 的非线性状态空间建模，Chen 和 Tsay (1993a) 的泛函系数自回归模型，Chen 和 Tsay (1993b) 的非线性可加自回归模型以及 Lewis 和 Stevens (1991)的多元适应回归样条. 这些扩展的基本思想或者用模拟方法来描述 x_t 条件分布的演变或者用数据驱动(data-driven)方法来探索一个序列的非线性特征. 最后，非参数和半参数方法，如核回归和人工神经网络，已经用于探索时间序列的非线性性. 在4.1节中，我们讨论一些可应用到金融时间序列分析上的非线性模型，包括一些非参数和半参数方法.

127

除了各种非线性模型的发展，人们还对能区分线性序列和非线性序列的检验统计量的研究感兴趣. 现可用的检验方法中既有参数检验又有非参数检验. 大多数参数检验是利用拉格朗日乘子法或似然比统计量. 非参数检验是依赖于 x_t 的高阶谱或针对混沌时间序列发展起来的维数相关系数. 我们在4.2节介绍一些非线性性检验方法. 4.3节和4.4节讨论非线性模型的建模与预测. 最后，在4.5节中给出非线性模型的应用.

4.1 非线性模型

统计文献中发展起来的大多数非线性模型都是针对式(4.3)中的均值方程的，参见 Priestley (1988) 和 Tong (1990) 中关于非线性模型的综述. 这里，我们的目的是介绍一些能用到金融时间序列上的非线性模型.

4.1.1 双线性模型

式(4.1)的线性模型只是式(4.2)函数 $f(.)$的一阶泰勒序列展开. 这样，向非线性的一个自然推广使用这个展开的二阶项来改进逼近. 这就是双线性模型的基本想法. 双线性模型可以定义为

$$x_t = c + \sum_{i=1}^{p}\phi_i x_{t-i} - \sum_{j=1}^{q}\theta_j a_{t-j} + \sum_{i=1}^{m}\sum_{j=1}^{s}\beta_{ij}x_{t-i}a_{t-j} + a_t \tag{4.4}$$

其中 p，q，m 和 s 是非负整数. 这个模型由 Granger 和 Andersen(1978)提出，并得到广泛研究. Subba Rao 和 Gabr(1984)讨论了这个模型的一些性质和应用，Liu 和 Brockwell(1988)研究了一般的双线性模型. 双线性模型的性质(如平稳性条件)通常通过两个步骤导出：(a)把模型写成状态空间形式，(b)利用状态转移方程把状态表示成过去的新息与随机系数向量的乘积.

式(4.4)中的双线性模型的特殊推广形式是有条件异方差性. 例如，考虑模型

$$x_t = \mu + \sum_{i=1}^{s} \beta_i a_{t-i} a_t + a_t \tag{4.5}$$

其中$\{a_t\}$是一个白噪声序列. x_t 的头两阶条件矩为

$$E(x_t \mid F_{t-1}) = \mu, \quad \mathrm{Var}(x_t \mid F_{t-1}) = \left(1 + \sum_{i=1}^{s} \beta_i a_{t-i}\right)^2 \sigma_a^2$$

这与第 3 章中的随机系数自回归（RCA）模型或条件异方差 ARMA（CHARMA）模型的头两阶矩相似.

128

例 4.1 考虑 CRSP 相等加权指数从 1926 年 1 月至 1997 年 12 月的月简单收益率，共 864 个观察值. 记此序列为 R_t，R_t 的样本偏自相关函数(PACF)在间隔为 1 和 3 点是显著局部自相关的，而 R_t^2 的样本 PACF 显示条件异方差可能依赖于过去的三个新息. 因此，我们对此序列采用特殊的双线性模型：

$$R_t = \mu + \phi_1 R_{t-1} + \phi_3 R_{t-3} + (1 + \beta_1 a_{t-1} + \beta_2 a_{t-2} + \beta_3 a_{t-3}) a_t.$$

假定 a_t 的条件分布为正态分布，我们用条件最大似然法得到所拟合的模型

$$R_t = 0.014 + 0.160 R_{t-1} - 0.104 R_{t-3} + (1 + 0.337 a_{t-1} - 0.022 a_{t-2} - 0.601 a_{t-3}) a_t \tag{4.6}$$

其中$\hat{\sigma}_a^2 = 0.005\,2$，各参数估计的标准差依次分别为 0.003，0.026，0.018，0.083，0.084 和 0.079. 只有 a_{t-2}的系数的估计值是不显著的. 定义此模型的残差序列为

$$\hat{a}_t = \frac{R_t - 0.014 - 0.160 R_{t-1} + 0.014 R_{t-3}}{1 + 0.337\,\hat{a}_{t-1} - 0.022\,\hat{a}_{t-2} - 0.601\,\hat{a}_{t-3}},$$

其中，对 $t \leqslant 3$，$\hat{a}_t = 0$. $\hat{a}_t$ 的样本 ACF 显示没有显著的前后相关性，但是平方序列$\hat{a}_t^2$ 有显著的前后相关性，这说明$\tilde{a}_t$ 不是独立的序列. 模型(4.6)的有效性需要进一步研究，为了做比较，我们再对此月简单收益率序列考虑一个 ARCH(3)模型，得到

$$\begin{aligned} R_t &= 0.013 + 0.222 R_{t-1} - 0.140 R_{t-3} + a_t, \\ \sigma_t^2 &= 0.002 + 0.168 a_{t-1}^2 + 0.000\,01 a_{t-2}^2 + 0.274 a_{t-3}^2 \end{aligned} \tag{4.7}$$

其中除 a_{t-2}^2的系数外，其他所有估计值都是高度显著的. 标准化的残差序列和它的平方序列都没有显著的前后相关性，这表示所建的 ARCH(3)模型对数据是合适的. 模型(4.6)和模型(4.7)有些相似，但后者似乎能更好地拟合数据.

4.1.2 门限自回归模型

此模型出现的动因是实际中经常观察到的几种非线性特征，例如一个过程在上升和下降阶段的非对称性. 它用分段线性模型来得到条件均值方程的更好逼近. 然而，与传统的分段线性模型不同的是：传统的模型是允许模型的变化发生在时间空间上，门限自回归 TAR 模型则是利用门限空间来改进线性逼近. 我们从简单的两个体制 AR(1)模型开始

129

$$x_t = \begin{cases} -1.5 x_{t-1} + a_t, & \text{若 } x_{t-1} < 0 \\ 0.5 x_{t-1} + a_t, & \text{若 } x_{t-1} \geqslant 0 \end{cases} \tag{4.8}$$

其中 a_t 是独立同分布的、都服从 $N(0, 1)$. 这里，时滞是 1，门限是 0. 图 4-1 所示的是 x_t 的 200 个模拟观察值的时间图. 图中有一条恒为 0 的水平线，这个图反映出 TAR 模型的几个特征. 第一，尽管在第一个体制中系数是 −1.5，过程还是几何遍历(geometrically ergodic)和平稳的. 事实上，模型(4.8)是几何遍历的充分必要条件是 $\phi_1^{(1)}<1$，$\phi_1^{(2)}<1$ 且 $\phi_1^{(1)}\phi_1^{(2)}<1$，其中 $\phi_1^{(i)}$ 是体制 i 的 AR 系数，参见 Petruccelli 和 Woolford(1984 年)，Chen 和 Tsay(1991 年). 遍历性在时间序列分析中是一个重要的概念. 例如，证明 x_t 的样本均值 $\bar{x}=\sum_{t=1}^{T} x_t/T$ 收敛于 x_t 的均值的统计理论称为遍历定理(ergodic theorem)，它被认为是和 iid 情形的中心极限定理同等重要的理论. 第二，该序列显示出非对称的上升和下降态势：若 x_{t-1} 是负的，则 x_t 将转向一个正值，由于这时有负的膨胀系数 −1.5. 然而，若 x_{t-1} 是正的，则经过几个时间指标后 x_t 序列变为负值. 从而，x_t 的时间图显示出体制 2 比体制 1 有更多的观察值，并且当序列取负值时它包含有很大的向上跳跃. 因此，此序列不是时间可逆的. 第三，此模型没有常数项，但 $E(x_t)$ 不为 0. 图中所示的这个具体实现的样本均值是 0.61，标准差为 0.07. 一般地，$E(x_t)$ 是两个体制下的条件均值的加权平均，通常不是 0，而每个体制的权重就是在平稳分布下 x_t 处于该体制的概率. 要使 TAR 模型的均值为 0，需要在有的体制中有非零常数项. 这一点与平稳线性模型有很大不同. 在平稳线性模型中非零常数项的存在意味着 x_t 的均值不为零.

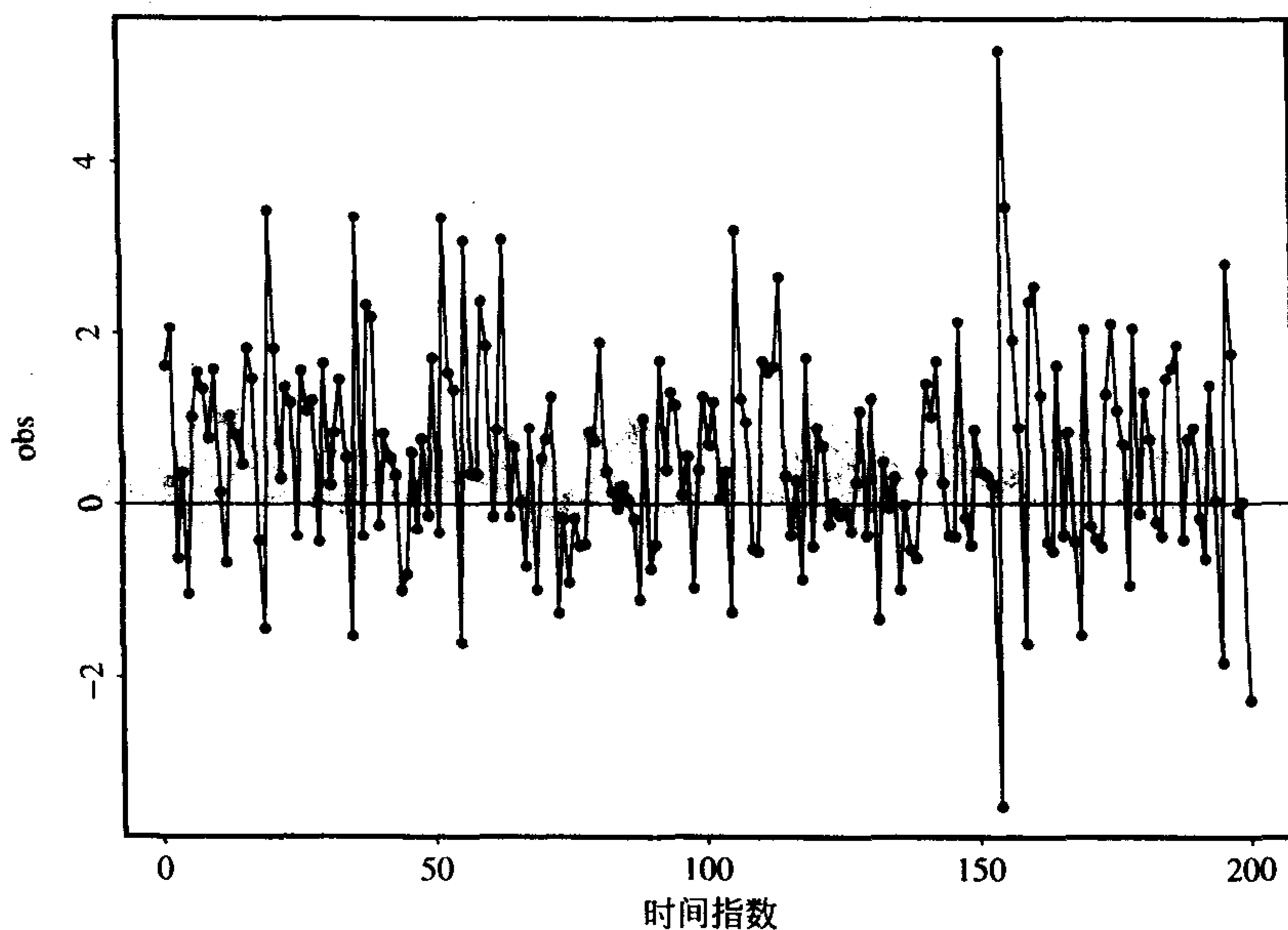


图 4-1 模拟的两个体制 TAR(1)序列的时间图

称一个时间序列 x_t 服从 k 个体制的自激发(self-exciting) TAR(SETAR)模型，如果它满足

$$x_t=\phi_0^{(j)}+\phi_1^{(j)}x_{t-1}-\cdots-\phi_p^{(j)}x_{t-p}+a_t^{(j)}, \quad \text{当 } \gamma_{j-1}\leqslant x_{t-d}<\gamma_j \text{ 时} \tag{4.9}$$

其中 k 和 d 是正整数，$j=1, \cdots, k$，γ_j 是满足 $-\infty=\gamma_0<\gamma_1<\cdots<\gamma_{k-1}<\gamma_k=\infty$ 的实数，上角标 (j) 用来表示体制，$\{a_t^{(j)}\}$ 是均值为 0、方差为 σ_j^2 的 iid 序列，且对不同的 j 是相互独立的，

x_{t-d}称为门限变量. 参数 d 称为延迟参数，γ_j 称为门限. 这里，对不同的体制，AR 模型是不同的，否则体制的个数可减少. 式(4.9)的自激发模型是在门限空间中分段线性的 AR 模型，它与回归分析中通常的分段线性模型的思想是相似的，只是模型的变化是以取得观察值的时间为顺序的. 只要 $k>1$，SETAR 模型就是非线性的.

一般 SETAR 的性质是难以得到的，一些特殊情形的性质可以在 Tong (1990)，Chan (1993)，Chan 和 Tsay(1998) 以及这几个文献的参考文献中看到. 近年来，人们对 TAR 模型和它们的应用的兴趣在增加，参见 Hansen (1997)，Tsay (1998) 和 Montgomery 等(1998). Tsay (1989) 提出了一个对一元 SETAR 检验和建模的方法. 式(4.9)的 SETAR 模型可以推广到只要求门限变量 z_t 为 F_{t-1} 可测的(即，它是 F_{t-1} 中元素的函数)情形，这时还要求 z_t 是平稳的，有在实直线的紧子集上连续的分布函数，并且 z_{t-d} 在 t 时刻是已知的. 这样的推广模型称为开环(open-loop)TAR 模型.

SETAR 模型在金融中的一个重要应用是处理波动率对正、负收益率的不对称响应. 此模型也用来研究指数期货和现金兑换价格中的套利交易，见第 8 章中关于多元时间序列的讨论. 这里我们只把注意力放在波动率建模上.

130 ~ 131

例 4.2 为了说明 TAR 模型在金融中的应用，我们考虑 IBM 股票的日对数收益率，从 1962 年 7 月 3 日到 1999 年 12 月 31 日共 9442 个观察值，以百分比形式给出并已包含分红. 图 4-2 所示的是此序列的时间图. 可见，在后期的波动率要大一些. 如果应用第 3 章中 GARCH 模型，对此序列我们得到如下的 AR(2)-GARCH(1, 1)模型

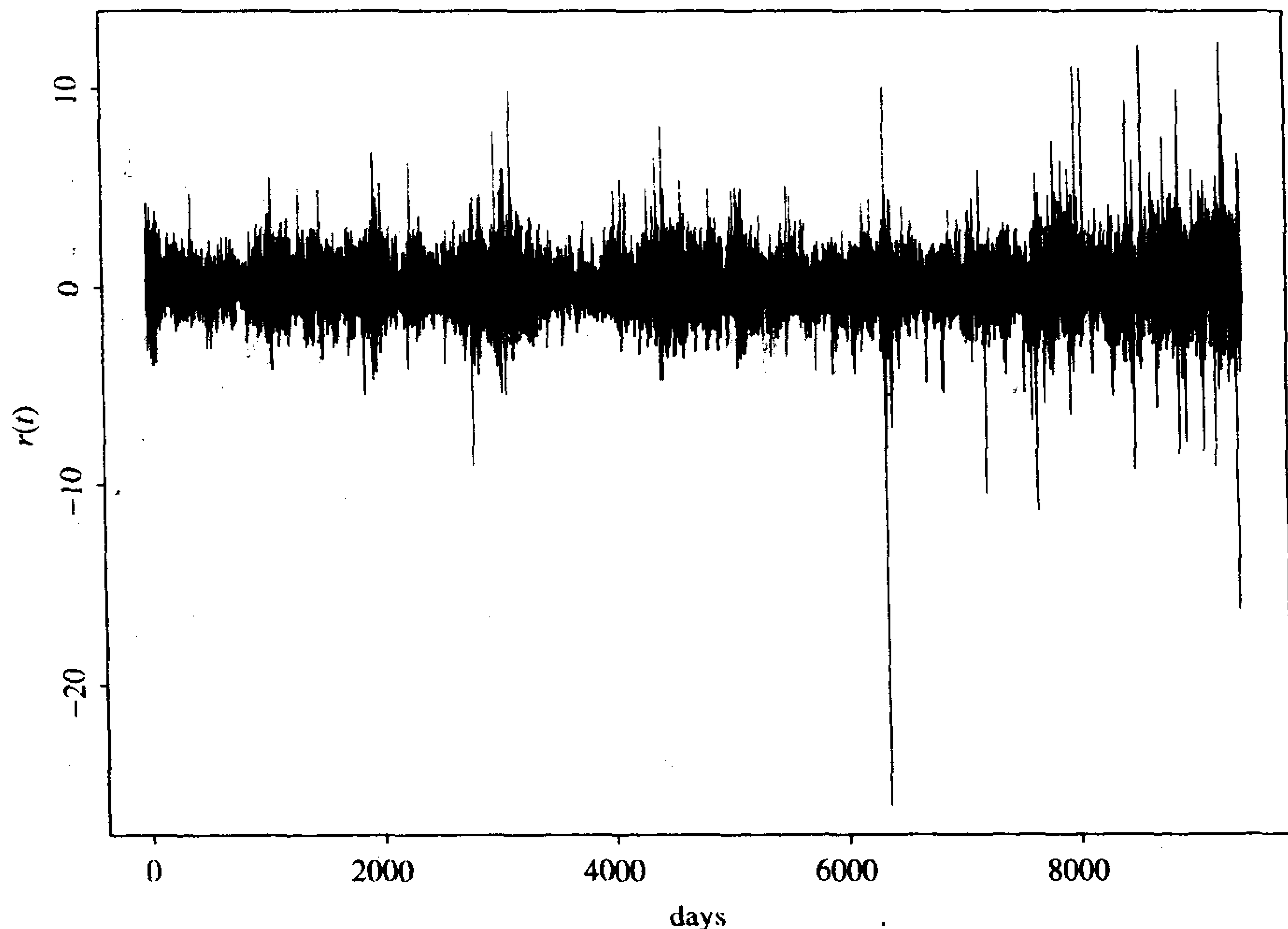


图 4-2 IBM 股票从 1962 年 7 月 3 日至 1999 年 12 月 31 日的日对数收益率的时间图

$$r_t = 0.067 - 0.023r_{t-2} + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 0.031 + 0.076a_{t-1}^2 + 0.915\sigma_{t-1}^2 \tag{4.10}$$

其中 r_t 是对数收益率，$\{\varepsilon_t\}$是均值为 0、方差为 1.0 的高斯白噪声序列. 均值方程中参数的标准差分别是 0.013 和 0.011，波动率方程中的标准差分别是 0.003、0.002 和 0.003. 除了 r_{t-2} 的系数外，其他估计都是高度显著的. 标准化残差的 Ljung-Box 统计量 $Q(10)=11.31(0.33)$、$Q(20)=27.00(0.14)$，其中括号里的数是 p 值. 对标准化残差的平方，$Q(10)=11.86(0.29)$、$Q(20)=19.19(0.51)$. 模型(4.10)在刻画数据的序列相关性和条件异方差性方面是合适的. 但由模型 r_t 得到的无条件均值为 0.065，要比样本均值 0.045 大一些，这显示该模型可能错了. 考虑到波动率对抖动 a_{t-1} 的正负号的不对称响应，可用 TAR 模型来改进此模型. 具体地，我们对该序列用一个 AR(2)-TAR-GARCH(1，1)模型，得到 [132]

$$r_t = 0.043 - 0.022r_{t-2} + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 0.002 + 0.097a_{t-1}^2 + 0.954\sigma_{t-1}^2 + (0.056 - 0.051a_{t-1}^2 - 0.067\sigma_{t-1}^2)I(a_{t-1} > 0)$$

其中

$$I(a_{t-1} > 0) = \begin{cases} 1, \text{当 } a_{t-1} > 0 \text{ 时}, \\ 0, \text{其他}. \end{cases}$$

因为在 5%的显著水平下波动率方程中估计值 0.002 是不显著的，我们把模型再改进为

$$r_t = 0.043 - 0.022r_{t-2} + a_t, \quad a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 0.098a_{t-1}^2 + 0.954\sigma_{t-1}^2 + (0.060 - 0.052a_{t-1}^2 - 0.069\sigma_{t-1}^2)I(a_{t-1} > 0) \tag{4.11}$$

其中均值方程中两个参数的标准差为 0.013 和 0.010，而 TAR-GARCH(1，1)模型中的各个参数的标准差分别为 0.003、0.004、0.005、0.004 和 0.009. 所有估计值在 5%水平是显著的. 由模型(4.11)得 r_t 的无条件均值为 0.042，这与 r_t 的样本均值非常接近. 基于 Ljung-Box 统计量的残差分析在标准化的残差中没有发现显著的序列相关性和条件异方差性. 均值方程中的 AR 系数很小，说明 IBM 股票的日对数收益率如预期的一样只有很弱的序列相关性. 然而，收益率的波动率方程表明了新息过程$\{a_t\}$中有强烈的相依性和条件方差中的不对称性. 把 TAR-GARCH(1，1)方程写成

$$\sigma_t^2 = \begin{cases} 0.098a_{t-1}^2 + 0.954\sigma_{t-1}^2, & \text{若 } a_{t-1} \leqslant 0, \\ 0.060 + 0.046a_{t-1}^2 + 0.885\sigma_{t-1}^2, & \text{若 } a_{t-1} > 0. \end{cases} \tag{4.12}$$

我们会看出一些有趣的推论. 首先，如果把 a_{t-1} 解释成 IBM 日对数收益率与它的条件期望的偏差，则当偏差非正时，波动率服从一个不带漂移项的 IGARCH(1，1)模型. 其次，当偏差为正时，波动率有一个持续参数 0.046+0.885=0.931，这靠近 1 但小于 1. 因而，当偏差为正时，波动率服从一个 GARCH(1，1)模型. 因此，波动率对正的和负的抖动的反映不同. 文献中还提出了其他的门限波动率模型(如 Rabemananjara 和 Zakoian(1993)，Zakoian (1994)).

注释：估计 AR(2)-Tar-GARCH(1，1)模型的 RATS 程序将在本章附录中给出. [133]

4.1.3 平滑转移 AR 模型

对 SETAR 模型的一种批评是它的条件均值方程不是连续的. 门限$\{\gamma_j\}$是条件均值函数 μ_t 的不连续点. 鉴于这种批评, 人们提出了平滑的 TAR 模型, 见 Chan 和 Tong (1986)、Teräsvirta(1994)以及这两篇论文的参考文献. 如果时间序列$\{x_t\}$满足

$$x_t = c_0 + \sum_{i=1}^{p} \phi_{0,i} x_{t-i} + F\left(\frac{x_{t-d} - \Delta}{s}\right)\left(c_1 + \sum_{i=1}^{p} \phi_{1,i} x_{t-i}\right) + a_t \tag{4.13}$$

则称$\{x_t\}$服从两个体制的 STAR(p)模型, 其中 d 是延迟参数, Δ 和 s 是模型的平移参数和尺度参数, $F(.)$是一个平滑的转移函数, 满足 $0 \leqslant F(.) \leqslant 1$. 在实际中, $F(.)$经常假定为三种形式之一: Logistic 函数、指数函数或者一个累积分布函数. 由式(4.13), STAR 模型的条件均值为如下两个方程的加权线性组合:

$$\mu_{1t} = c_0 + \sum_{i=1}^{p} \phi_{0,i} x_{t-i},$$

$$\mu_{2t} = (c_0 + c_1) + \sum_{i=1}^{p} (\phi_{0,i} + \phi_{1,i}) x_{t-i}.$$

权重由 $F\left(\frac{x_{t-d} - \Delta}{s}\right)$以一个连续的方式决定. 上述两个方程也决定了 STAR 模型的性质. 例如 STAR 模型平稳性的必要条件是这两个 AR 多项式的零点都在单位圆外. STAR 的优点是条件均值函数是可微的, 但经验表明其中的参数 Δ 和 s 是难以估计的. 特别地, 大多数实证研究表明 Δ 和 s 估计的标准差相当大, t 比大致为 1, 见 Teräsvirta(1994). 这种不确定性会导致在解释所估计出的模型时复杂性较大.

例 4.3 我们用 3M (Minnesota Mining and Manufacturing)公司股票从 1946 年 2 月到 1997 年 12 月的月简单收益率来解释 STAR 模型的应用. 若考虑一个 ARCH 模型, 我们得到如下的 ARCH(2):

$$R_t = 0.014 + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 0.003 + 0.108 a_{t-1}^2 + 0.151 a_{t-2}^2, \tag{4.14}$$

其中各估计值的标准差分别为 0.002、0.0003、0.045 和 0.058. 如前面讨论的, 这样的 ARCH 模型不能描述股票波动率对正负抖动的不对称反应. 用 STAR 模型可以克服这一困难. 把 STAR 模型应用于 3M 股票的月收益率, 我们得到模型

134

$$R_t = 0.017 + a_t, \quad a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = (0.002 + 0.256 a_{t-1}^2 + 0.141 a_{t-2}^2) + \frac{0.002 - 0.314 a_{t-1}^2}{1 + \exp(-1\,000 a_{t-1})}, \tag{4.15}$$

其中均值方程中常数项的标准差为 0.002, 而波动率方程中的各估计值的标准差分别为 0.000 3、0.092、0.056、0.001 和 0.102. 为了简化估计, logistic 转移函数的尺度参数 1000 是事先给定的. 这个 STAR 模型给对正负抖动反应的不对称性提供了支持. 对绝对值很大的负 a_{t-1}, 波动率模型近似为 ARCH(2)模型:

$$\sigma_t^2 = 0.002 + 0.256 a_{t-1}^2 + 0.141 a_{t-2}^2.$$

对很大的正 a_{t-1}，波动率过程近似如下 ARCH(2)模型：

$$\sigma_t^2 = 0.005 - 0.058a_{t-1}^2 + 0.141a_{t-2}^2.$$

此模型中 a_{t-1}^2 的系数为负，这与直观相违，但绝对值很小. 事实上，对较大的正抖动 a_{t-1}，ARCH 效应较弱，即使参数估计仍然是统计显著的. 计算时所用的 RATS 程序将在附录中给出.

4.1.4 马尔可夫转换模型

在非线性时间序列分析中使用概率转移的思想在 Tong(1983) 中讨论. 用类似的想法，但强调一个经济量在各个状态之间的非周期转移，Hamilton (1989) 考虑马尔可夫转换自回归(MSA)模型. 这里转移是由一个隐含的两个状态马尔可夫链驱动的. 如果时间序列 x_t 满足

$$x_t = \begin{cases} c_1 + \sum_{i=1}^{p} \phi_{1,i} x_{t-i} + a_{1t}, & \text{如果 } s_t = 1, \\ c_2 + \sum_{i=1}^{p} \phi_{2,i} x_{t-i} + a_{2t}, & \text{如果 } s_t = 2, \end{cases} \tag{4.16}$$

其中 s_t 是在{1，2}中取值的马尔可夫链，转移概率为

$$P(s_t = 2 \mid s_{t-1} = 1) = w_1, \quad P(s_t = 1 \mid s_{t-1} = 2) = w_2,$$

那么时间序列 x_t 服从 MSA 模型. 新息序列 $\{a_{1t}\}$、$\{a_{2t}\}$ 都是零均值、有有限方差的独立同分布随机变量序列，并且两序列之间是相互独立的. 较小 w_i 意味着模型在状态 i 上停留更长的时间. 事实上，$1/w_i$ 是过程停留在状态 i 上的期望持续时间长度. 由定义可见，MSA 模型利用一个隐含的马尔可夫链掌握从一个条件均值函数到另一个条件均值函数的转移. 这与 SETAR 模型不同，在那里转移是由一个具体的延迟变量来决定的. 总之，SETAR 模型以一个确定的方案掌握模型的转移，而 MSA 是以一个随机的方式来掌握模型的转移. 在实际中，由于状态的随机性，在 MSA 模型中不能确定 x_t 是属于哪一个状态. 当样本容量很大时，可利用一些滤波方法对 x_t 的状态作出推断. 然而对 SETAR 模型，只要 x_{t-d} 被观察到了，x_t 的体制就已知了. 这个差别在预测时有重要的实际涵义. 例如，MSA 模型的预测总是由各个状态下的子模型产生的预测的线性组合，但对 SETAR 模型，一旦 x_{t-d} 被观察到，其预测就是单个体制下的预测，如果预测步长超过延迟 d，SETAR 的预测也是各个体制下预测的线性组合. MSA 模型的估计要比其他模型困难得多，因为状态不是可直接观察的. Hamilton(1990)采用 EM 算法估计 MSA 模型，EM 算法是一种在取期望和取最大值之间重复迭代的统计方法. McCulloch 和 Tsay(1994)考虑用马尔可夫链-蒙特卡罗(MCMC)方法估计一般的 MSA 模型. 我们在第 10 章将讨论 MCMC 方法.

135

McCulloch 和 Tsay(1993) 把式(4.16)的 MSA 模型推广为转移概率 w_1 和 w_2 是在 $t-1$ 时刻可获得的一些解释变量的 logistic 或概率的函数情形. Chen，McCulloch 和 Tsay (1997)用马尔可夫转换的思想作为工具进行对非嵌套的非线性时间序列模型的比较和选取(例如，比较双线性模型和 SETAR 模型)，每一个候选模型用一个状态来表示. 这种选择模型的方法是贝叶

斯分析中常用的机会比(odds ratio)方法的推广. 另外，MSA 模型可以容易地推广到多个状态(多于两个状态)的情形，但计算量将迅速增加. 对计量经济学中马尔可夫转换模型更多的讨论，请参见 Hamilton (1994)的第 22 章.

例 4.4 考虑美国的实际国民总产值(GNP)的季度增长率，时间段是 1947 年第二季度到 1991 年的第一季度，以百分比形式给出. 图 4-3 是经过季度调节的数据. 可见大多数增长率是正的，其中有一条水平的零增长线. 这个序列在经济时间序列的非线性分析中广泛应用，Tiao 和 Tsay (1994)，Potter (1995)用 TAR 模型，Hamilton (1989)，McCulloch 和 Tsay (1994) 用马尔可夫转换模型都研究过此序列.

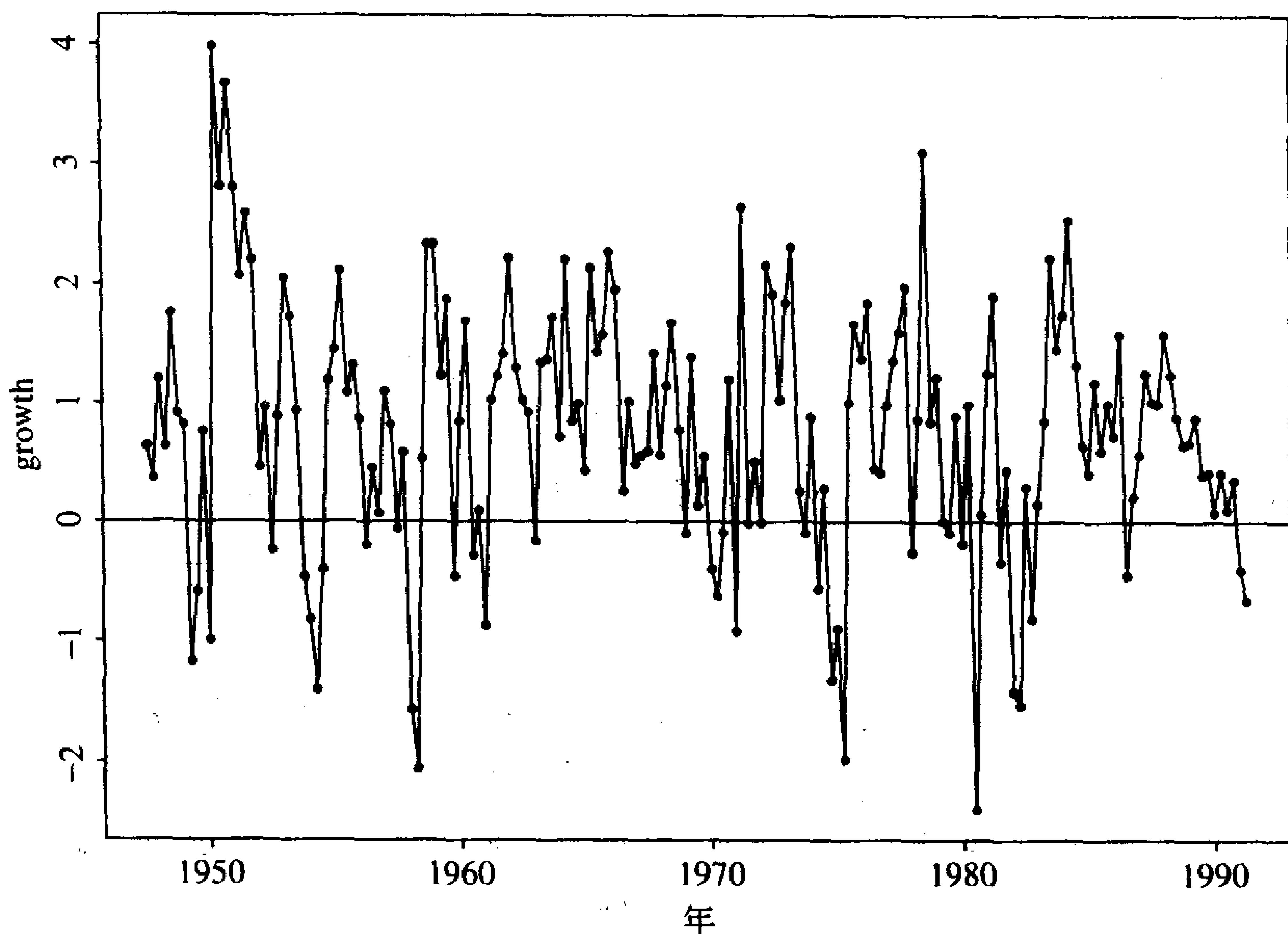


图 4-3 美国实际 GNP 季度增长率的时间图，时间段为 1947 年第 2 季度至 1991 年第 1 季度，数据经过季度性调节并以百分数形式给出

采用当 $p=4$ 时的式(4.16)MSA 模型，用 MCMC 方法(第 10 章中将讨论)，McCulloch 和 Tsay(1994)得到的估计列在表 4-1 中. 结果中有几点有趣的发现：第一，状态 1 下模型的平均增长率为 $0.909/(1-0.265-0.029+0.126+0.11)=0.965$，状态 2 下模型的平均增长率为 $-0.42/(1-0.216-0.628+0.073+0.097)=-1.288$. 这样，状态 1 对应有正增长率的季度或膨胀期，状态 2 对应有负增长率的季度或紧缩期. 第二，状态 2 时，参数估计的相对较大的标准差反应出在这个状态的观察值较少，这一点可从图 4-3 上看出，较少的季度具有负增长率. 第三，对不同的状态，转移概率会不同，走出紧缩期的概率较大(0.286 对 0.118). 第四，把 $1/w_i$ 当作停留在状态 i 的平均持续时间长度，我们看到紧缩期和膨胀期的平均长度分别大约是 3.69 和 11.31 季度. 这样，平均起来，美国经济的紧缩大致持续一年，而膨胀大致持续三年. 最后，两个状态下 x_{t-2}的 AR 系数的估计值差别很大，显示出美国经济的动态规律在紧缩期和膨胀期是不同的.

表 4-1

参数	c_i	ϕ_1	ϕ_2	ϕ_3	ϕ_4	σ_i	w_i
				状态 1			
估计值	0.909	0.265	0.029	−0.126	−0.110	0.816	0.118
标准差	0.202	0.113	0.126	0.103	0.109	0.125	0.053
				状态 2			
估计值	−0.420	0.216	0.628	−0.073	−0.097	1.017	0.286
标准差	0.324	0.347	0.377	0.364	0.404	0.293	0.064

注：对美国实际 GNP 的季度增长率（经季度调节）运用 $p=4$ 时的 MSA 模型的估计结果．估计值和它们的标准差是 5 000 次迭代的 Gibbs 抽样的后验均值和标准差．

4.1.5 非参数方法

在一些金融应用中，我们可能没有足够的信息来确定两个变量 Y 和 X 之间的非线性结构．我们希望利用计算设备和计算方法上的进展来探索 Y 与 X 之间的函数关系．这就要用到非参数的方法和技术．然而，非参数方法的运用不是没有代价的，它们是高度地依赖于数据并容易导致过度拟合(overfitting)．这里的目的是要介绍金融应用中的一些非参数方法和利用非参数方法的一些非线性模型．所讨论的非参数方法包括核回归、局部最小二乘估计和神经网络．

非参数的本质是平滑．考虑两个金融变量 Y 和 X，满足

$$Y_t = m(X_t) + a_t \tag{4.17}$$

其中 $m(.)$ 是光滑的但未知的任意函数，$\{a_t\}$ 是白噪声序列．我们想由数据估计非线性函数 $m(.)$．为了简单起见，考虑 $X=x$ 这个特定时候，$m(.)$ 的估计问题，也就是 $m(x)$ 的估计问题．假定对应 $X=x$，我们有重复独立的观察 y_1，…，y_T，则数据变为

$$y_t = m(x) + a_t, \qquad t = 1,\cdots,T$$

对数据取平均得到：

$$\frac{\sum_{t=1}^{T} y_t}{T} = m(x) + \frac{\sum_{t=1}^{T} a_t}{T}$$

由大数定律，当 T 趋于∞时，抖动的平均趋于 0，故 $\bar{y} = \sum_{t=1}^{T} y_t / T$ 是 $m(x)$ 的一个相合估计．

在金融时间序列中，在 $X=x$ 点，我们不能有重复的观察．所观察到的是 $\{(y_t, x_t)\}$ 对 $t=1$，…，T．如果函数 $m(.)$ 是充分光滑的，对应于 $X_t \approx x$ 的那些 Y_t 的值仍然可以提供 $m(x)$ 的较精确的近似．而对应于远离 x 的 X_t 的那些 Y_t 值就不能提供对 $m(x)$ 的较好近似．作为折中，人们用 y_t 的加权平均来代替简单的平均来估计 $m(x)$．权重应满足：对应于 x 附近的 X_t 的那些 Y_t 所给权重较大，而对应于远离 x 的 X_t 的 Y_t 所给权重较小．写成数学式子，对给定 x，$m(x)$ 的估计为

136~138

$$\hat{m}(x) = \frac{1}{T}\sum_{t=1}^{T} w_t(x) y_t \tag{4.18}$$

其中权重 $w_t(x)$满足：对应于 x 附近的 x_t 的 y_t 所给 $w_t(x)$较大，对应于远离 x 的 x_t 的 y_t 所给的 $w_t(x)$较小.

由式(4.18)，$\hat{m}(x)$就是一个局部加权平均，其中权重由两个因素决定. 第一个因素是距离度量(即 x_t 与 x 之间的距离)，第二个因素是对给定的距离权重的分配. x_t 与 x 间不同的距离度量方法和不同的权重分配产生不同的非参数方法. 下面讨论通常使用核回归和局部线性回归方法.

4.1.5.1 核回归

核回归也许是最常用的平滑非参数回归方法. 这时的权重是由一个核函数决定，典型的核函数是概率密度函数，一般记为 $K(x)$，满足

$$K(x) \geqslant 0, \quad \int K(z)dz = 1$$

为了增强在距离度量上的灵活性，人们往往用一个变量 $h>0$ 来对核函数重新尺度化

$$K_h(x) = \frac{1}{h}K(x/h), \quad \int K_h(z)dz = 1 \tag{4.19}$$

这里的 h 称为带宽(bandwidth). 权函数定义为

$$w_t(x) = \frac{K_h(x - x_t)}{\frac{1}{T}\sum_{t=1}^{T} K_h(x - x_t)}, \tag{4.20}$$

其中分母是一个标准化常数，它使得平滑能适应于变量 X 的局部强度，而且权重之和为 T. 把式(4.20)代入平滑公式(4.18)中，我们得到著名的 Nadaraya-Watson 核估计：

139

$$\hat{m}(x) = \frac{1}{T}\sum_{t=1}^{T} w_t(x) y_t = \frac{\sum_{t=1}^{T} K_h(x - x_t) y_t}{\sum_{t=1}^{T} K_h(x - x_t)}, \tag{4.21}$$

参见 Nadaraya(1964)和 Watson(1964). 在实际中，有很多核函数 $K(x)$可供选择. 然而，从理论和实际应用的两个角度考虑，只有几种核函数是常用的，其中包括高斯核函数

$$K_h(x) = \frac{1}{h\sqrt{2\pi}}\exp\left(-\frac{x^2}{2h^2}\right)$$

和 Epanechnikov 核函数(Epanechnikov(1969))

$$K_h(x) = \frac{0.75}{h}\left(1 - \frac{x^2}{h^2}\right)I\left(\left|\frac{x}{h}\right| \leqslant 1\right),$$

其中 $I(A)$是一个示性函数：若 A 成立则 $I(A)=1$，否则 $I(A)=0$. 图 4-4 所示的是当 $h=1$ 时，高斯核和 Epanechnikov 核的图像.

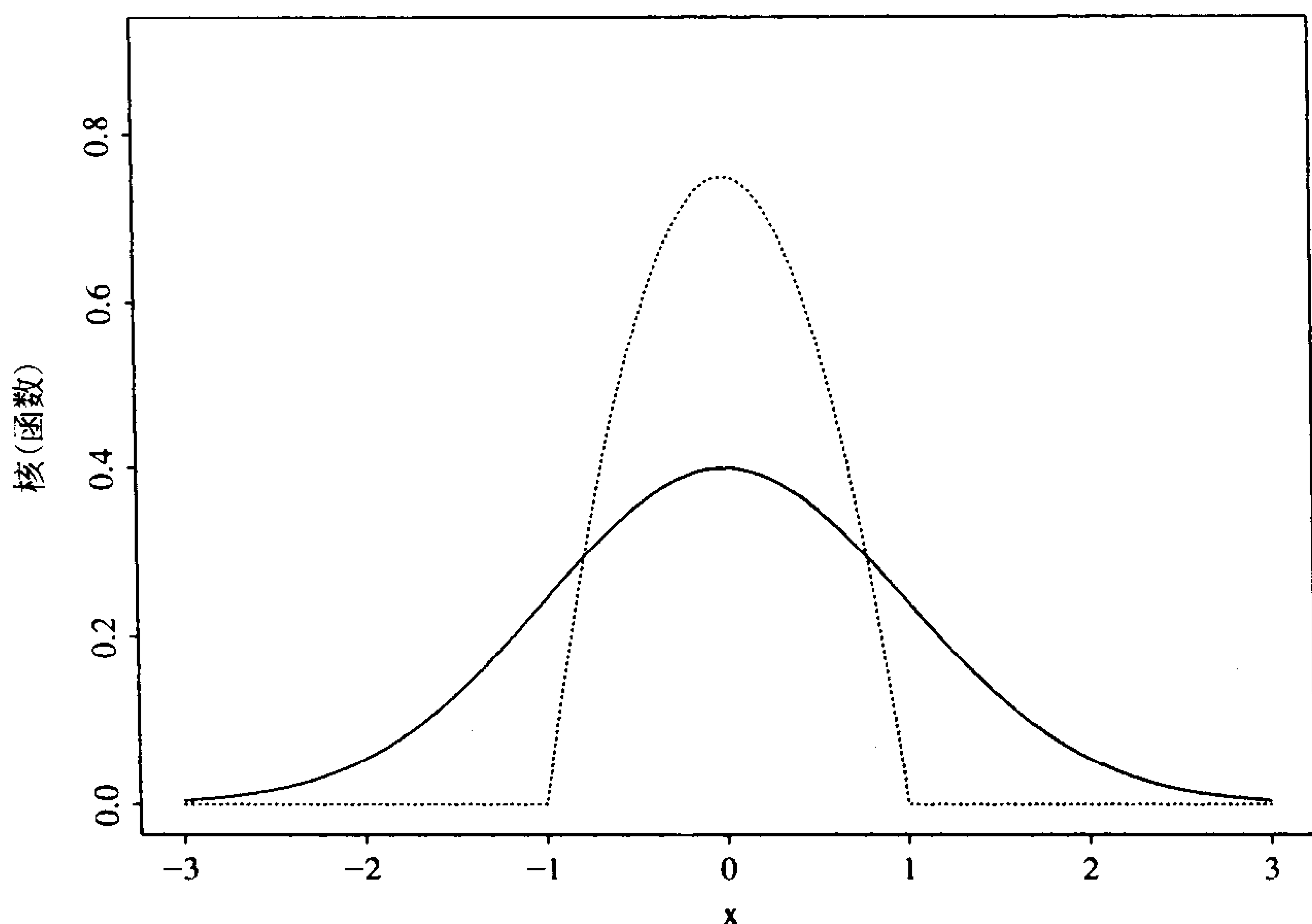


图 4-4 标准高斯核(实线)和 Epanechnikov 核(虚线)，带宽均为 $h=1$

为了理解带宽 h 所起的作用，我们来评价一些带 Epanechnikov 核的 Nadaraya-Watson 估计在观察值$\{x_t\}$上的表现. 考虑两个极端情形：第一，若$h\to 0$,则

$$\hat{m}(x_t)\to\frac{K_h(0)y_t}{K_h(0)}=y_t,$$

140

这表明很小的 h，会重现数据本身；第二，若 $h\to\infty$，则

$$\hat{m}(x_t)\to\frac{\sum_{t=1}^{T}K_h(0)y_t}{\sum_{t=1}^{T}K_h(0)}=\frac{1}{T}\sum_{t=1}^{T}y_t=\bar{y}$$

说明很大的 h 会导致过度平滑曲线——样本均值. 一般地，带宽 h 的作用如下：当 h 很小时，权重只集中在 x_t 的邻域内的少数观察值上，当 h 很大时，权重会分散到 x_t 的较大邻域内. 总之，h 的选择在核回归分析中起着重要作用. 这就是核回归中有名的带宽的选择问题.

4.1.5.2 带宽的选择

带宽的选择有好几种方法，参见 Härdle(1990). 第一种方法是塞入法，它基于核平滑函数的均方误差(MSE)的渐近展式. 考虑

$$E[\hat{m}(x)-m(x)]^2$$

其中 $m(.)$是真实函数，在一些正则性条件下，我们可通过最小化 MSE 得到最优带宽. 最优带宽一般要依赖于待估的未知量，必须由数据用某种初次平滑去估计这些未知量. 要得到合理的最优带宽估计，需要多步迭代. 在实际应用中，初次平滑的选择是一个问题.

第二种方法是留一交叉验证(cross-validation). 首先，去除一个观察值(x_j, y_j). 用剩下

的 $T-1$ 个数据点得到如下的在 x_j 点的平滑：

$$\hat{m}_{h,j}(x_j) = \frac{1}{T-1}\sum_{t\neq j} w_t(x_j) y_t$$

它是 y_j 的一个估计值. 第二，对 $j=1, \cdots, T$ 都做第一步，定义函数

$$CV(h) = \frac{1}{T}\sum_{j=1}^{T}[y_j - \hat{m}_{h,j}(x_j)]^2 W(x_j)$$

其中 $W(.)$是一个非负的权重函数，如果必要，可用来降低对边界点上的权重. 因为对接近边界的点，通常只有很少数的数据点与之相邻，所以有降低它们权重的必要. 函数 $CV(h)$称为交叉验证函数，因为它验证了平滑函数预测$\{y_t\}_{t=1}^{T}$的能力. 可以通过最小化 $CV(.)$函数来选择带宽 h.

141

4.1.5.3 局部线性回归法

假定式(4.17)中的 $m(.)$的二阶导数存在，并且在给定的属于 $m(.)$支撑的点 x 处连续. 记数据为$\{(y_t, x_t)\}_{t=1}^{T}$. 非参数回归中的局部线性回归法要找出 a 和 b，使下式达最小值：

$$L(a,b) = \sum_{t=1}^{T}[y_t - a - b(x-x_t)]^2 K_h(x-x_t), \tag{4.22}$$

其中 $K_h(.)$是式(4.19)中的核函数，h 是带宽. 记 a 的估计结果为$\hat{a}$，则 $m(x)$的估计定义为$\hat{a}$，x 为一个观察数据.

在最小二乘理论下，式(4.22)是一个加权最小二乘问题，我们可以推导出 a 的显式解. 具体地，求 $L(a, b)$对 a 和 b 的偏导数，令偏导数等于 0，我们得到有两个未知数的两个方程：

$$\sum_{t=1}^{T} K_h(x-x_t) y_t = a\sum_{t=1}^{T} K_h(x-x_t) + b\sum_{t=1}^{T}(x-x_t)K_h(x-x_t),$$

$$\sum_{t=1}^{T} y_t(x-x_t)K_h(x-x_t) = a\sum_{t=1}^{T}(x-x_t)K_h(x-x_t) + b\sum_{t=1}^{T}(x-x_t)^2 K_h(x-x_t).$$

定义

$$s_{T,l} = \sum_{t=1}^{T} K_h(x-x_t)(x-x_t)^l, \qquad l = 0,1,2.$$

上述两个方程组变为

$$\begin{bmatrix} s_{T,0} & s_{T,1} \\ s_{T,1} & s_{T,2} \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} K_h(x-x_t) y_t \\ \sum_{t=1}^{T}(x-x_t)K_h(x-x_t) y_t \end{bmatrix}$$

从而，

$$\hat{a} = \frac{s_{T,2}\sum_{t=1}^{T} K_h(x-x_t) y_t - s_{T,1}\sum_{t=1}^{T}(x-x_t)K_h(x-x_t) y_t}{s_{T,0}s_{T,2} - s_{T,1}^2}.$$

142

上式的分子和分母可进一步简化

$$s_{T,2}\sum_{t=1}^{T}K_h(x-x_t)y_t - s_{T,1}\sum_{t=1}^{T}(x-x_t)K_h(x-x_t)y_t$$

$$=\sum_{t=1}^{T}[K_h(x-x_t)(s_{T,2}-(x-x_t)s_{T,1})]y_t,$$

$$s_{T,0}s_{T,2}-s_{T,1}^2=\sum_{t=1}^{T}K_h(x-x_t)s_{T,2}-\sum_{t=1}^{T}(x-x_t)K_h(x-x_t)s_{T,1}$$

$$=\sum_{t=1}^{T}K_h(x-x_t)[s_{T,2}-(x-x_t)s_{T,1}]$$

总之，我们有

$$\hat{a}=\frac{\sum_{t=1}^{T}w_t y_t}{\sum_{t=1}^{T}w_t} \tag{4.23}$$

其中 w_t 定义为

$$w_t=K_h(x-x_t)[s_{T,2}-(x-x_t)s_{T,1}]$$

在实际应用中，为了避免分母为 0 的情况，我们使用 $m(x)$的如下估计：

$$\hat{m}(x)=\frac{\sum_{t=1}^{T}w_t y_t}{\sum_{t=1}^{T}w_t+\frac{1}{T^2}} \tag{4.24}$$

注意到式(4.24)有一个很好的特性，就是权重 w_t 满足

$$\sum_{t=1}^{T}(x-x_t)w_t=0.$$

另外，如果只假定式(4.17)中 $m(.)$有一阶导数，通过求

$$\sum_{t=1}^{T}(y_t-a)^2K_h(x-x_t)$$

的最小值点，可得到前面介绍过的 Nadaraya-Watson 估计. 一般地，如果假定$m(x)$有有界的 k 阶导数，则可用一个$(k-1)$次多项式代替式(4.22)中的线性多项式. 我们把式(4.24)中的估计称为局部线性回归平滑. Fan(1993)证明了在一些正则条件下局部线性回归有一些重要的抽样性质. 带宽的选择可以用与前面讨论过的相同的方法. [143]

4.1.5.4 时间序列的应用

在时间序列分析中，解释变量通常是序列的延迟值. 考虑单个解释变量的简单情形. 这时模型(4.17)变成

$$x_t=m(x_{t-1})+a_t$$

核回归和局部线性回归方法可直接应用. 当有多个解释变量存在时，需要作一些修改后再用非参数方法. 对核回归，我们可以用多元核函数，如具有给定的协方差矩阵的多元正态密度函数

$$K_h(\boldsymbol{x})=\frac{1}{(h\sqrt{2\pi})^p\mid\Sigma\mid^{1/2}}\exp\left(-\frac{1}{2h^2}\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}\right)$$

其中 p 是解释变量的个数，Σ是事先给定的正定矩阵. 也可以用一元核函数的乘积作为多元核函数，例如

$$K_h(\boldsymbol{x})=\prod_{i=1}^{p}\frac{0.75}{h_i}\left(1-\frac{x_i^2}{h_i^2}\right)I\left(\left|\frac{x_i}{h_i}\right|<1\right)$$

这一方法比较简单，但忽略了解释变量之间的关系.

下面介绍几个由非参数方法导出的非线性时间序列模型.

4.1.6 函数系数 AR 模型

非参数方法的新近成果使得研究人员在提出非线性模型时可以放松参数型的约束. 有些情形下，非参数方法用来对数据进行初步研究，以帮助选择一个参数型非线性模型. Chen 和 Tsay (1993a) 采用这种方法，提出函数系数自回归(functional-coefficient autoregressive, FAR)模型

$$x_t=f_1(X_{t-1})x_{t-1}+\cdots+f_p(X_{t-1})x_{t-p}+a_t \tag{4.25}$$

其中 $X_{t-1}=(x_{t-1},\ \cdots,\ x_{t-k})'$是由 x_t 的延迟值所构成的向量. 如果有必要，X_{t-1}可以包含在 $t-1$ 时刻的其他解释变量. 式(4.25)中的函数 $f_i(.)$假定是连续、甚至是二次可微的. 前面讨论过的绝大部分非线性模型都是 FAR 模型的特殊情形. 在实际应用中，可用非参数方法(如核回归或局部线性回归)来估计函数型系数 $f_i(.)$，特别是当 X_{t-1}的维数较低的时候可以这样做. 最近，Cai，Fan 和 Yao(1999) 用局部线性回归方法估计 $f_i(.)$，证明了用 FAR 模型可以在向前 1 步预测上获得很大改进.

144

4.1.7 非线性可加 AR 模型

把非参数方法应用到非线性时间序列分析中的一个主要困难是“维数灾难”(curse of dimensionality). 考虑一个一般的 AR(p)过程 $x_t=f(x_{t-1},\ \cdots,\ x_{t-p})+a_t$. 直接用非参数方法来估计 $f(.)$，将需要 p 维平滑. 当 p 很大时，这难以做到，尤其是在数据点的个数不是很大时. 克服这个困难的一个简单而有效的方法是考虑可加模型，这样的模型只需要低维的平滑. 若

$$x_t=f_0(t)+\sum_{i=1}^{p}f_i(x_{t-i})+a_t \tag{4.26}$$

其中 $f_i(.)$是几乎处处连续的函数，则称时间序列 x_t 服从非线性可加 AR(简记为 NAAR)模型. 因为每个 $f_i(.)$只有一个自变量，可以用一维平滑方法，从而避免维数灾难. 在应用中，在给定 $f_j(.)$，$j\neq i$ 的条件下，估计 $f_i(.)$，这种迭代方法可用来估计 NAAR 模型. 更多细节和 NAAR 模型的例子，参见 Chen 和 Tsay (1993b).

可加性假设是限制性相当强的，在应用中要仔细检验. Chen，Liu 和 Tsay(1995)考虑了检验可加性的检验统计量.

4.1.8 非线性状态空间模型

利用 MCMC 方法的最新进展(Gelfand 和 Smith，1990)，Carlin，Polson 和 Stoffer (1992) 提出了非线性状态空间建模的蒙特卡罗方法. 所考虑的模型为

$$S_t = f_t(S_{t-1}) + u_t, \quad x_t = g_t(S_t) + v_t \tag{4.27}$$

其中 S_t 是状态向量，$f_t(.)$和 $g_t(.)$是依赖某些未知参数的已知函数，$\{u_t\}$是具有零均值、非负定的协方差矩阵Σ_u 的独立同分布多元随机向量序列，$\{v_t\}$是均值为零、方差为 σ_v^2 的 iid 随机变量序列，$\{u_t\}$与$\{v_t\}$独立. 因为对一个非线性系统，需要给定 S_{t-1}的条件下 S_t 的整个条件分布，蒙特卡罗方法可用来处理状态转移方程的非线性演变. Kitagawa(1998) 及其参考文献中考虑了其他的对非线性时间序列分析的数值平滑方法. MCMC 方法(或者一些高强度计算的数值方法)是非线性时间序列分析的有力工具，但这些方法的潜力还没有充分发掘出来. 然而，式(4.27)模型中假定 $f_t(.)$和 $g_t(.)$的形式已知，这是在实际应用中运用上述方法的障碍. 克服这一局限的方案是用在对 FAR 和 NAAR 模型分析时所用的非参数方法，在用非线性状态空间模型之前，对 $f_t(.)$和 $g_t(.)$进行具体化.

145

4.1.9 神经网络

现代数据分析的一个常见论题就是神经网络. 神经网络方法可归类为半参数方法. 神经网络的有关文献非常多，它在很多学科中有应用，应用成功的程度也不尽相同，参见 Ripley (1993)的第 2 节所列的应用，和第 10 节中关于在金融中应用的评注. Cheng 和 Titterington (1994)提供了从统计学角度看神经网络的信息. 在这一小节中，我们主要关注前馈(feed-forward)神经网络，它将原始信息输入到由一个或几个神经元(或称为结点)构成的输入层，而这些结点向前连接到下一层，直到到达输出层. 图 4-5 举了一个简单的例子，它是带有一个隐层的、用来处理一元时间序列分析的简单的前馈神经网络. 其输入层有两个结点，隐层有三个结点. 输入层的每个结点都向前连接到每一个隐层的结点上，而这些隐层的结点都连接到输出层的一个单独结点. 我们称这个网络为一个 2-3-1 的前馈神经网络. 更复杂的神经网络包括那些带回馈连接的网络都已经在文献中提出，但与我们学习相关最多的还是前馈神经网络.

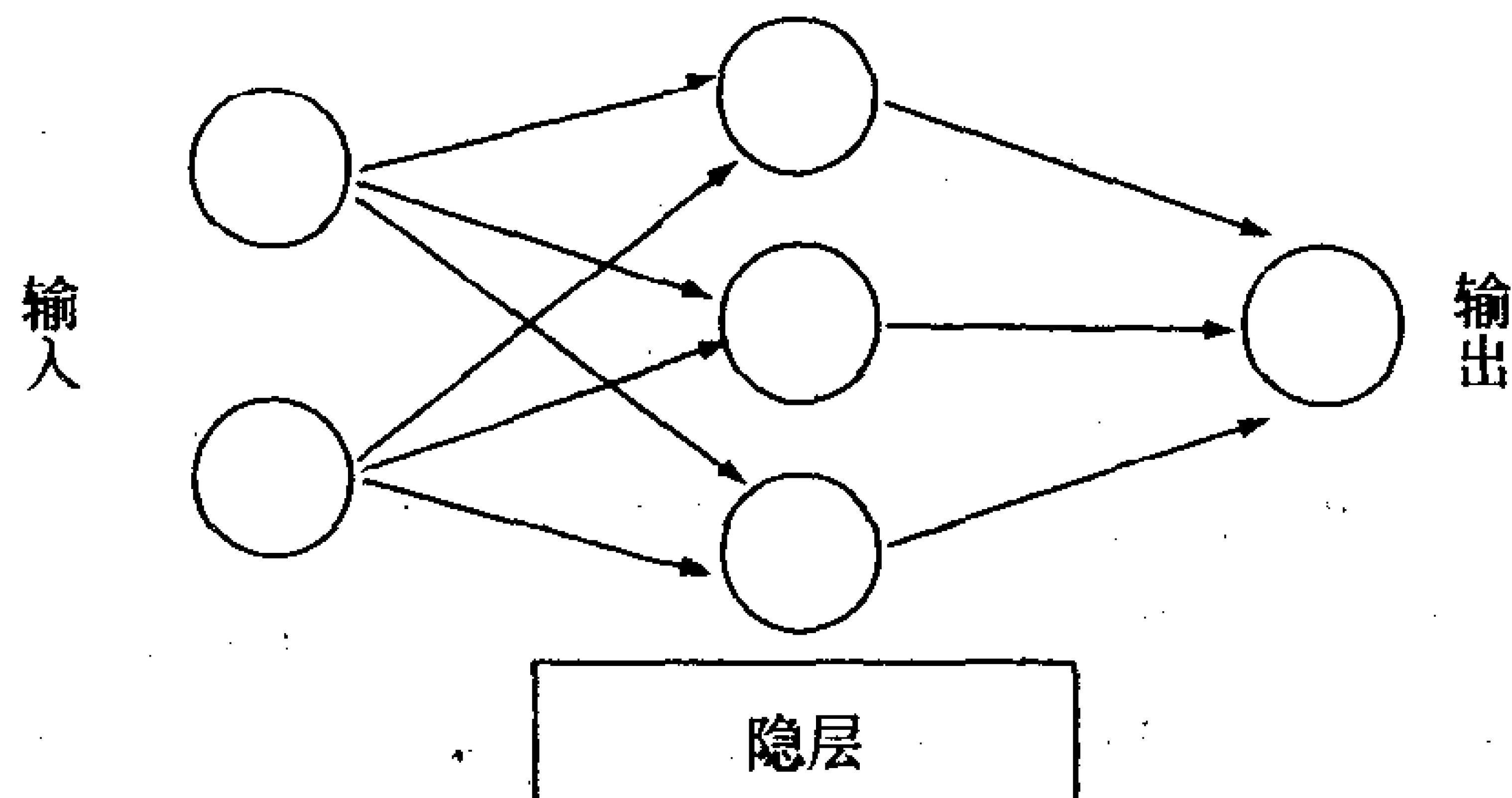


图 4-5 一个用于处理一元时间序列分析的带有一个隐层的前馈神经网络

146

4.1.9.1 前馈神经网络

神经网络由一层向下一层处理加工信息通过一个“激活函数”(activation function)来完成.考虑带有一个隐层的前馈神经网络，隐层的第 j 个结点定义为：

$$h_j = f_j\Big(\alpha_{0j} + \sum_{i\to j} w_{ij}x_i\Big) \tag{4.28}$$

其中 x_i 是输入层第 i 个结点的值，$f_j(.)$是一个激活函数，典型的是取 logistic 函数：

$$f_j(z) = \frac{\exp(z)}{1+\exp(z)}$$

α_{0j}称为偏差，对 $i\to j$ 求和是指对输入层中所有指向 j 的结点求和，w_{ij}是权重. 比如在图 4-5 中的 2-3-1 前馈神经网络的隐层中的第 j 个结点：

$$h_j = \frac{\exp(\alpha_{0j} + w_{1j}x_1 + w_{2j}x_2)}{1+\exp(\alpha_{0j} + w_{1j}x_1 + w_{2j}x_2)}, \qquad j = 1,2,3. \tag{4.29}$$

输出层的结点定义为：

$$o = f_o\Big(\alpha_{0o} + \sum_{j\to o} w_{jo}h_j\Big) \tag{4.30}$$

其中激活函数 $f_o(.)$是线性函数或 Heaviside 函数. 如果 $f_o(.)$是线性函数，则

$$o = \alpha_{0o} + \sum_{j=1}^{k} w_{jo}h_j$$

其中 k 是隐层中结点数目. 在一个 Heaviside 函数中，我们定义 $f_o(z)=1$ 如果 $z>0$，否则 $f_o(z)=0$. 把带有 Heaviside 函数的神经元称为门限神经元，“1”代表神经元发出信息. 例如，考虑图 4-5 中的 2-3-1 网络，如果激活函数是线性的，则其输出值是

$$o = \alpha_{0o} + w_{1o}h_1 + w_{2o}h_2 + w_{3o}h_3$$

如果 $f_o(.)$是一个 Heaviside 函数，则输出值是

$$o = \begin{cases} 1, & \text{如果 } \alpha_{0o} + w_{1o}h_1 + w_{2o}h_2 + w_{3o}h_3 > 0 \\ 0, & \text{如果 } \alpha_{0o} + w_{1o}h_1 + w_{2o}h_2 + w_{3o}h_3 \leqslant 0 \end{cases}$$

[147]

联合各个层来看，前馈神经网络的输出值可以写成

$$o = f_o\Big[\alpha_{0o} + \sum_{j\to o} w_{jo} f_j(\alpha_{0j} + \sum_{i\to j} w_{ij}x_i)\Big] \tag{4.31}$$

如果允许输入层直接连接到输出层，则该网络变成

$$o = f_o\Big[\alpha_{0o} + \sum_{i\to o}\alpha_{io}x_i + \sum_{j\to o} w_{jo} f_j(\alpha_{0j} + \sum_{i\to j} w_{ij}x_i)\Big] \tag{4.32}$$

其中第一个求和号是对输入结点求和. 当输出层的激活函数是线性的，输入层到输出层的直接连接意味着输出值是输入值的一个线性函数. 从而，在这个特殊情形下，式(4.32)是线性模型的一般形式. 对于图 4-5 的 2-3-1 网络，如果输出激活函数是线性的，则式(4.31)变成

$$o = \alpha_{0o} + \sum_{j=1}^{3} w_{jo}h_j$$

其中 h_j 在式(4.29)给出. 从而，该网络有 13 个参数. 如果用式(4.32)，则网络变为：

$$o = \alpha_{0o} + \sum_{i=1}^{2}\alpha_{io}x_i + \sum_{j=1}^{3}w_{jo}h_j$$

其中 h_j 在式(4.29)给出．该网络参数的数目增加到 15 个．

我们称等式(4.31)或式(4.32)中的函数是半参数函数，是因为它的函数形式是已知的，而结点个数、结点的偏差和权重是未知的．式(4.32)中从输入层到输出层的直接连接意味着该网络跳过了隐层．我们把这样的网络称为一个跳过隐层的前馈网络．

前馈网络在神经网络的文献中认为是多层感知器．它们在紧集上可以通过增加隐层的结点数目一致逼近任何连续函数．见 Hornik，Stinchcombe 和 White (1989)；Hornik (1993)；Chen 和 Chen (1995)．神经网络的性质是多层感知器的通用估计性质．简言之，带有一个隐层的前馈神经网络可以看成是一种对一般的连续非线性函数参数化的方法． 148

4.1.9.2 训练和预测

神经网络的应用包括两个步骤．第一个步骤是训练网络(例如建立一个网络，包括确定结点数目以及估计它们的偏差和权重)．第二个步骤是推断，特别是预测．在训练阶段，数据通常被分成两部分不相重叠的子样本．第一组子样本用来估计已知的前馈神经网络的参数．在第二组子样本中用这样建立的网络来进行预测和计算它的预测精度．通过比较预测效果，可以选出比其他效果好的“最优”网络来进行推断．这正是在统计模型选择中应用非常广泛的交叉验证法的思想．其他的模型选择方法也是可以利用的．

在一个时间序列应用中，令$\{(r_t, \boldsymbol{x}_t) \mid t=1, \cdots, T\}$为网络训练可利用的数据．其中 $\boldsymbol{x}_t$ 表示输入值的向量，而 r_t 是所关心的序列(比如一项资产的对数收益率)．对于一个给定的网络，令 o_t 为输入值是 $\boldsymbol{x}_t$ 的网络的输出值；见等式(4.32)．训练一个神经网络相当于选择它的偏差和权重，最小化某种合适的标准，比如，最小二乘

$$S^2 = \sum_{t=1}^{T}(r_t - o_t)^2.$$

这是一个非线性估计问题，可以用一些迭代方法解决．为了保证所拟合函数的光滑性，对前面的最小化问题加上一些附加的约束．在神经网络文献中，向后传播(Back Propagation (BP))学习算法是一种网络训练的普遍方法．BP 方法是由 Bryson 和 Ho(1969)提出的，它从输出层开始逆向进行，并且反复用梯度规则来修正偏差和权重．Ripley (1993)中的附录 2A 提供了向后传播的推导．一旦一个前馈神经网络建立，它就能在预测子样本中用来计算预测了．

例 4.5 为举例说明金融中神经网络的应用，我们考虑 IBM 股票从 1926 年 1 月到 1999 年 12 月的月对数收益率，以百分比形式并且包括分红．我们把数据分为两组子样本．第一组子样本是从 1926 年 1 月到 1997 年 12 月的 864 个观测数据，用来建立模型．用式(4.32)的带有三个输入值和含两个结点的隐层构成的模型，我们可以得到一个 3-2-1 的序列网络．这三个输入值分别记为 r_{t-1}，r_{t-2}和 r_{t-3}，偏差和权重由下式给出：

$$\hat{r}_t = 3.22 - 1.81f_1(\boldsymbol{r}_{t-1}) - 2.28f_2(\boldsymbol{r}_{t-1}) - 0.09r_{t-1} - 0.05r_{t-2} - 0.12r_{t-3}, \quad (4.33)$$

其中 $\boldsymbol{r}_{t-1} = (r_{t-1}, r_{t-2}, r_{t-3})$，两个 logistic 函数是

$$f_1(\boldsymbol{r}_{t-1}) = \frac{\exp(-8.34 - 18.97r_{t-1} + 2.17r_{t-2} - 19.17r_{t-3})}{1 + \exp(-8.34 - 18.97r_{t-1} + 2.17r_{t-2} - 19.17r_{t-3})},$$ 149

$$f_2(\boldsymbol{r}_{t-1}) = \frac{\exp(39.25 - 22.17r_{t-1} - 17.34r_{t-2} - 5.98r_{t-3})}{1 + \exp(39.25 - 22.17r_{t-1} - 17.34r_{t-2} - 5.98r_{t-3})}$$

前面模型的残差标准差是 6.56．作为对照，我们也用同一数据建立 AR 模型，得到

$$r_t = 1.101 + 0.077r_{t-1} + a_t, \quad \sigma_a = 6.61 \tag{4.34}$$

这个模型的残差标准差比式(4.33)的前馈神经网络的稍微大一点．

预测比较

1998 年和 1999 年 IBM 股票的月对数收益率构成了第二组子样本，用来评价神经网络对样本以外的数据的预测效果．作为比较的基准模型，我们用第一组子样本的样本均值 r_t 对第 2 组子样本的所有月收益率进行 1 步向前预测，这相当于假定 IBM 股票的月对数价格服从一个带漂移项的随机游动模型．这个基准模型均方预测误差(MSE)是 91.85．式(4.34)的 AR(1)模型的 1 步向前预测的 MSE 是 91.70．这样看，AR(1)模型比基准模型效果稍好一点．而式(4.33)的 3-2-1 前馈网络的 MSE 是 91.74，与 AR(1)模型效果基本相同．

注释：我们用初始权重设为缺省值的 S-Plus 程序来完成对前馈网络的估计．要了解更多信息见 Venables 和 Ripley (1999)．我们有限的经验表明估计结果是变化的．对例 4.5 中用到的 IBM 股票收益率，3-2-1 网络的样本外数据的 MSE 最低可为 89.46，最高可到 93.65．如果我们改变隐层的结点数目，MSE 的变动范围可以更宽．附录 B 给出了例 4.5 中用到的 S-Plus 指令．

例 4.6 前馈网络的优良特性包括其灵活性和广泛适用性．举个例子，我们用一个网络，它的输出层的激活函数是 Heaviside 函数，预测例 4.5 中考虑的 IBM 股票的价格变动方向．定义一个方向函数

$$d_t = \begin{cases} 1, & \text{如果 } r_t \geqslant 0 \\ 0, & \text{如果 } r_t < 0 \end{cases}$$

我们用 8 个输入结点和 4 个隐层结点建立一个 8-4-1 网络来预测第一组子样本的 d_t，其 8 个输入结点是由前 4 个 r_t 和 d_t 的延迟值组成的．这个网络用于计算 1 步向前的向上运动(例如正收益)概率在第二组子样品中的后续月份．图 4-6 表示的是一个典型的概率预测的输出值和第二组子样本的实际方向，后者在图中以“o”标出．在图中添加一条概率为 0.5 的水平线．如果我们采用较为严格的方法，当概率预测结果大于或等于 0.5，令 $\hat{d}_t=1$ 否则令 $\hat{d}_t=0$．那么神经网络的预测成功率为 0.58．当从一个估计变成另一个时，这个网络的成功率会显著改变，这个网络有 49 个参数．为了更深入理解，我们做一个模拟研究，将 8-4-1 网络运行 500 次，计算用以前的方法预测时向上或向下运动出错数目．在这 500 次中，平均错误的数目和错误的中值分别是 11.28 和 11，错误最多和最少的数目分别是 18 和 4．作为对照，我们再用带漂移项的随机游动模型进行 500 次模拟预测，即

$$\hat{d}_t = \begin{cases} 1, & \text{如果 } \hat{r}_t = 1.19 + \epsilon_t \geqslant 0 \\ 0, & \text{否则} \end{cases}$$

其中 1.19 是 IBM 股票从 1926 年 1 月到 1997 年 12 月的平均月对数收益率，$\{\epsilon_t\}$ 是服从 $N(0,1)$ 的独立同分步随机变量序列．预测错误的数目的平均值和中位数变为 10.53 和 11，而错误最多和最少的数目分别是 17 和 5．图 4-7 是两种模拟方法的预测出错数目的柱状图．结果表明

150

对于IBM股票的月对数价格，8-4-1前馈神经网络不比假定为一个带漂移项的随机游动的简单模型更好.

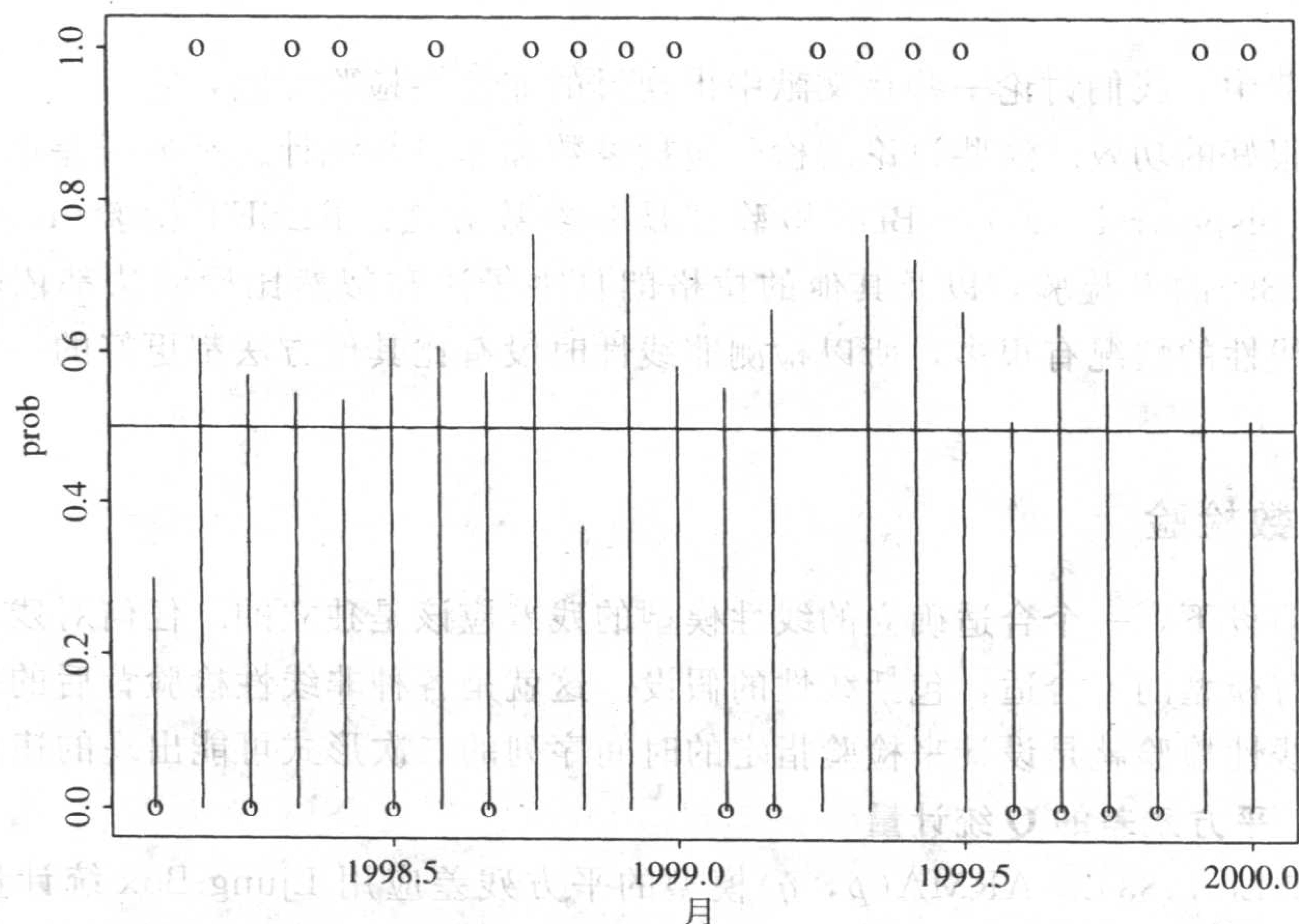


图 4-6 用 8-4-1 前馈神经网络对 IBM 股票正的月收益进行 1 步向前概率预测. 预测期是从 1998 年 1 月到 1999 年 12 月

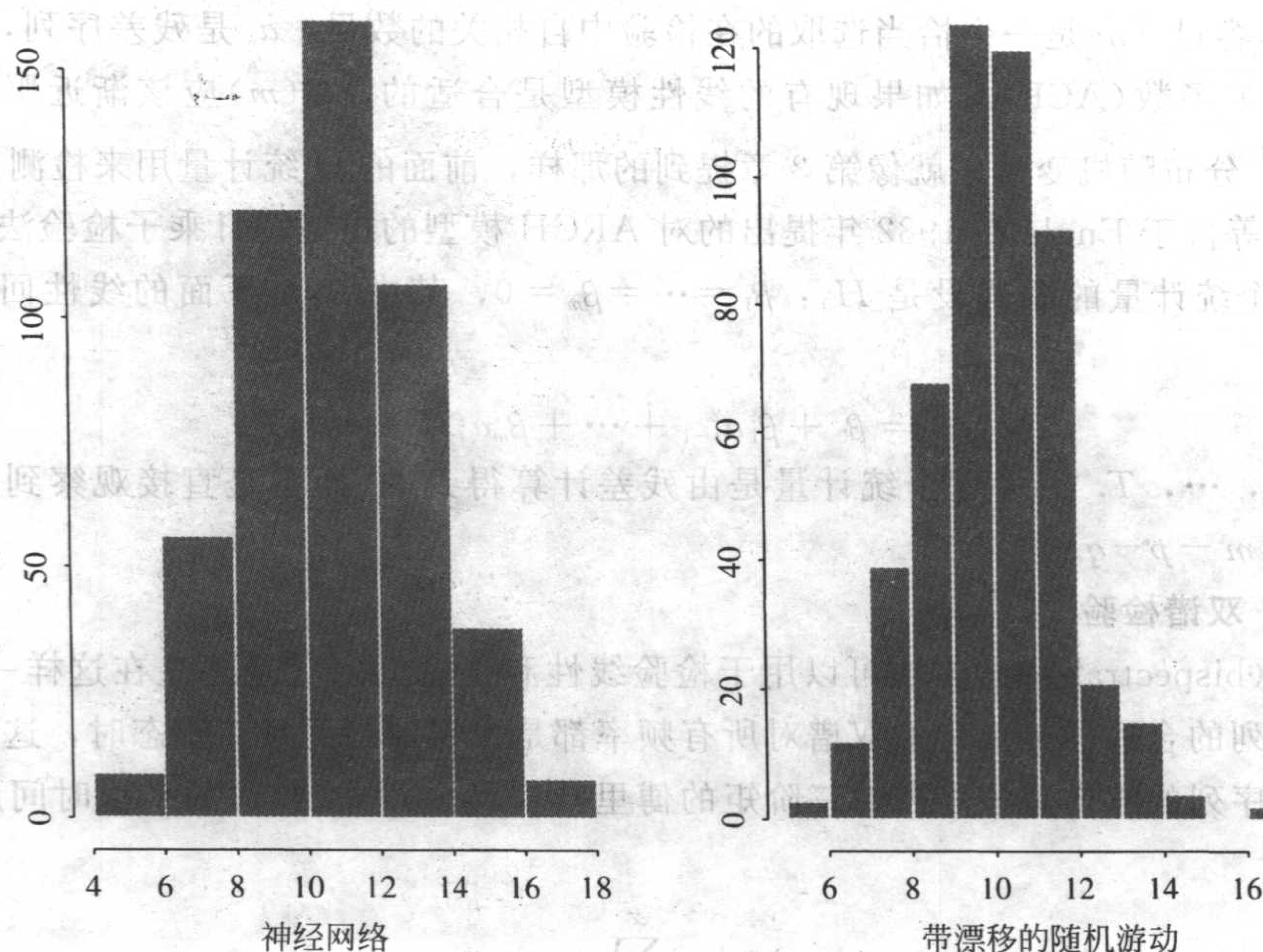


图 4-7 对 IBM 股票的月对数收益率的变动方向进行预测错误的数目的柱状图. 预测期是从 1998 年 1 月到 1999 年 12 月

4.2 非线性检验

在这一小节中，我们讨论一些在文献中出现过的非线性检验方法，它们对 4.1 节中考虑的非线性模型有很好的功效. 这些讨论的检验包括参数和非参数统计. 平方残差的 Ljung-Box 统计，双谱检验(bispectral test)和 BDS 检验都是非参数方法. RESET 检验(Ramsey, 1969), Tsay(1986, 1989)的 F 检验，以及其他的拉格朗日乘子法和似然比检验法都依赖特定的参数函数. 因为非线性的情况有很多，所以检测非线性时没有比其他方法都更好的一种单独的检验方法.

4.2.1 非参数检验

151~152

在线性原假设下，一个合适确立的线性模型的残差应该是独立的. 任何对残差独立性的违背都说明了现有模型的不合适，包括线性的假设. 这就是各种非线性检验背后的基本思想. 特别地，一些非线性检验就是设计来检验指定的时间序列的二次形式可能出现的违背情况.

4.2.1.1 平方残差的 Q 统计量

McLeod 和 Li(1983)对 ARMA(p, q)模型的平方残差应用 Ljung-Box 统计量来检查模型的不足. 这个检验统计量是

$$Q(m) = T(T+2)\sum_{i=1}^{m}\frac{\hat{\rho}_i^2(a_t^2)}{T-i}$$

其中 T 是样本容量，m 是一个恰当选取的在检验中自相关的数目，a_t 是残差序列，而$\hat{\rho}_i(a_t^2)$是 a_t^2 的 i 阶自相关函数(ACF). 如果现有的线性模型是合适的，$Q(m)$应该渐近服从自由度为 $m-p-q$的χ^2 分布随机变量. 就像第 3 章提到的那样，前面的 Q 统计量用来检测 a_t 的条件异方差性，渐近等价于 Engle 在 1982 年提出的对 ARCH 模型的拉格朗日乘子检验法统计量；见 3.3.3 节. 这个统计量的原假设是 H_0：$\beta_1=\cdots=\beta_m=0$，其中 β_i 是下面的线性回归的 a_{t-i}^2 的系数：

$$a_t^2 = \beta_0 + \beta_1 a_{t-1}^2 + \cdots + \beta_m a_{t-m}^2 + e_t$$

其中 $t=m+1$, …, T. 因为这个统计量是由残差计算得到的(而不是直接观察到的收益率)，所以自由度是 $m-p-q$.

4.2.1.2 双谱检验

双谱检验(bispectral test)方法可以用于检验线性和正态性. 它是建立在这样一个结果上，即线性时间序列的合适的标准化的双谱对所有频率都是恒定的，而且在正态时，这个定值就是零. 一个时间序列的双谱是该序列的三阶矩的傅里叶变换. 对式(4.1)的平稳时间序列 x_t，其三阶矩定义为

$$c(u,v) = g\sum_{k=-\infty}^{\infty}\psi_k\psi_{k+u}\psi_{k+v} \tag{4.35}$$

其中 u 和 v 是整数，$g=E(a_t^3)$，$\psi_0=1$，$\psi_k=0$ 当 $k<0$ 时. 对式(4.35)作傅里叶变换，我们

得到:

$$b_3(w_1,w_2) = \frac{g}{4\pi^2}\Gamma[-(w_1+w_2)]\Gamma(w_1)\Gamma(w_2) \tag{4.36}$$

其中 $\Gamma(w) = \sum_{u=0}^{\infty}\psi_u \exp(-iwu)$,而 $i=\sqrt{-1}$,w_i 是频率. x_t 的谱密度函数由下式给出: 153

$$p(w) = \frac{\sigma_a^2}{2\pi} \mid \Gamma(w) \mid^2,$$

其中 w 是频率. 从而,函数变为

$$b(w_1,w_2) = \frac{\mid b_3(w_1,w_2) \mid^2}{p(w_1)p(w_2)p(w_1+w_2)} = \text{恒定值,对所有}(w_1,w_2) \tag{4.37}$$

双谱检验利用式(4.37)的性质. 基本上,我们通过合适的选取格子点来估计式(4.37)中的函数 $b(w_1, w_2)$,并且用一种类似 Hotelling 的 T^2 统计量的检验统计量来检验 $b(w_1, w_2)$的恒定性. 对一个线性高斯序列,$E(a_t^3)=g=0$,所以双谱对所有频率(w_1, w_2)都恒为零. 对双谱检验的更详细讨论见 Priestley (1988),Subba Rao 和 Gabr (1984),Hinich(1982). 有限的经验表明当样本容量很大时,这种检验法有很好的功效.

4.2.1.3 BDS 统计量

Brock,Dechert 和 Scheinkman (1987)提出了一个检验统计量,通常称为 BDS 检验,用来检测一个时间序列的独立同分布假设. 这个统计量不同于其他讨论过的检验统计量,因为后者主要集中在 x_t 的 2 阶或 3 阶性质上. BDS 检验法的基本思想是利用混沌时间序列分析中常见的"相关性求和". 给定一个 k 维时间序列 X_t 和观察值$\{X_t\}_{t=1}^{T_k}$,定义其相关性求和为

$$C_k(\delta) = \lim_{T_k\to\infty}\frac{2}{T_k(T_k-1)}\sum_{i<j}I_\delta(X_i,X_j), \tag{4.38}$$

其中 $I_\delta(u, v)$是一个示性变量,它当$\|u-v\|<\delta$ 时等于 1,否则等于 0,其中$\|.\|$是上确界范数(supnorm). 相关性求和度量了相距不超过 δ 的一对数据$\{X_t\}$占所有数据对的比重. 考虑下一个时间序列 x_t. 设 k 维向量 $X_t^k=(x_t, x_{t+1}, \cdots, x_{t+k-1})'$,称为 k 历史. BDS 检验的思想如下,把一个 k 历史看成 k 维空间中的一个点. 如果$\{x_t\}_{t=1}^{T}$确实是独立同分布随机变量序列的话,那么 k 历史$\{X_t\}_{t=1}^{T_k}$应该在 k 维空间中看不出样式来. 从而,相关性求和应该满足关系 $C_k(\delta)=[C_1(\delta)]^k$. 如果违背了此式就表明 x_t 不是独立同分布的. 举一个简单而又说明问题的例子,考虑一列独立同分布的随机变量,服从[0, 1]上的均匀分布. 令$[a, b]$是[0, 1]的子区间,考虑 2 历史(x_t, x_{t+1}),它表示 2 维空间中的一个点. 在独立同分布假设下,在$[a, b]\times[a, b]$中 2 历史数目的期望应该等于$[a, b]$中 x_t 数目期望的平方. 这个想法可以用相关性求和对应的样本进行正式的检验. 定义 154

$$C_l(\delta,T) = \frac{2}{T_k(T_k-1)}\sum_{i<j}I_\delta(X_i^*,X_j^*), \qquad l=1,k,$$

其中 $T_l=T-l+1$,当 $l=1$ 时 $X_i^*=x_i$;当 $l=k$ 时 $X_i^*=X_i^k$. 在$\{x_t\}$是独立同分布的,有非退化的分布函数 $F(.)$的原假设下,Brock,Dechert,Scheinkman (1987) 指出

$$C_k(\delta,T) \to [C_1(\delta)]^k \quad (\text{以概率 }1,\text{当 } T\to\infty)$$

对任意固定的 k 和 δ. 更进一步，统计量$\sqrt{T}\{C_k(\delta,\ T)-[C_1(\delta,\ T)]^k\}$渐近于正态分布，它是零均值，方差是

$$\sigma_k^2(\delta) = 4\left[N^k + 2\sum_{j=1}^{k-1} N^{k-j}C^{2j} + (k-1)^2C^{2k} - k^2NC^{2k-2}\right]$$

其中 $C=\int[F(z+\delta)-F(z-\delta)]dF(z)$，$N=\int[F(z+\delta)-F(z-\delta)]^2dF(z)$. 注意到 $C_1(\delta,T)$ 是 C 的相合估计，而 N 的 一致估计可以为

$$N(\delta,T) = \frac{6}{T_k(T_k-1)(T_k-2)}\sum_{t<s<u} I_\delta(x_t,x_s)I_\delta(x_s,x_u)$$

于是 BDS 检验统计量定义为

$$D_k(\delta,T) = \sqrt{T}\{C_k(\delta,T) - [C_1(\delta,T)]^k\}/\sigma_k(\delta,T) \tag{4.39}$$

其中 $\sigma_k(\delta,\ T)$是用 $C_1(\delta,\ T)$和 $N(\delta,\ T)$分别代替 C 和 N 时通过$\sigma_k(\delta)$得到的. 这个检验统计量的极限分布是标准正态的. 应用 BDS 检验的更多的讨论和例子请见 Hsieh (1989)；Brock, Hsieh, LeBaron (1991). 在实际应用中，在用 BDS 检验前，如果数据有线性相依性，我们要去掉其相依性. 这种检验对 k 和 δ 的选择可能很敏感，尤其当 k 比较大的时候.

4.2.2 参数检验

将目光转向参数检验，我们考虑 Ramsey 的 RESET 检验(1969)以及它的推广. 我们仍然讨论一些检验统计量来检测门限非线性. 为简化记号起见，我们在讨论中使用向量和矩阵. 如果需要的话，读者可以在第 8 章的附录 A 简单复习一下向量和矩阵的相关内容.

4.2.2.1 RESET 检验

155 Ramsey(1969) 提出一种对线性最小二乘回归分析的规范检验. 这个检验称为 RESET 检验，可以容易地应用于线性 AR 模型. 考虑一个线性 AR(p)模型：

$$x_t = \boldsymbol{X}'_{t-1}\boldsymbol{\phi} + a_t \tag{4.40}$$

其中 $\boldsymbol{X}_{t-1}=(1,\ x_{t-1},\ \cdots,\ x_{t-p})'$，$\boldsymbol{\phi}=(\phi_0,\ \phi_1,\ \cdots,\ \phi_p)'$. RESET 检验的第一步是获得式(4.40)中的最小二乘估计$\hat{\boldsymbol{\phi}}$，然后计算拟合值$\hat{x}_t=\boldsymbol{X}'_{t-1}\hat{\boldsymbol{\phi}}$，残差$\hat{a}_t=x_t-\hat{x}_t$，以及残差平方和 $SSR_0=\sum_{t=p+1}^{T}\hat{a}_t^2$，其中 T 是样本容量. 第二步考虑线性回归：

$$\hat{a}_t = \boldsymbol{X}'_{t-1}\boldsymbol{\alpha}_1 + \boldsymbol{M}'_{t-1}\boldsymbol{\alpha}_2 + v_t \tag{4.41}$$

其中 $\boldsymbol{M}_{t-1}=(\hat{x}_t^2,\ \cdots,\ \hat{x}_t^{s+1})'$对某个 $s\geqslant1$，再计算最小二乘残差

$$\hat{v}_t = \hat{a}_t - \boldsymbol{X}'_{t-1}\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{M}'_{t-1}\hat{\boldsymbol{\alpha}}_2$$

以及回归的残差平方和 $SSR_1=\sum_{t=p+1}^{T}\hat{v}_t^2$. RESET 检验的基本思想是，如果式(4.40) 的AR(p) 模型合适，那么式(4.41)中的$\boldsymbol{\alpha}_1$和$\boldsymbol{\alpha}_2$应该为零. 这可以通过一般的 F 统计量进行检验，式(4.41)

的 F 统计量如下：

$$F=\frac{(SSR_0-SSR_1)/g}{SSR_1/(T-p-g)} \quad 其中\ g=s+p+1 \tag{4.42}$$

在线性和正态性假定下，它服从自由度是 g 和 $T-p-g$ 的 F 分布.

注释： 因为 $\hat{x}_t^k$(对 $k=2, \cdots, s+1$)与 $\boldsymbol{X}_{t-1}$ 以及它们自身之间趋向有高度相关性，所以用 $\boldsymbol{M}_{t-1}$ 的与 $\boldsymbol{X}_{t-1}$ 没有线性关系的主成分对式(4.41)进行拟合. 主成分分析是一种降维的统计工具，参见第 8 章的更多内容.

Keennan(1985)提出一种对时间序列的非线性检验方法，只用到 $\hat{x}_t^2$，并且修正 RESET 检验法的第二步来回避 $\hat{x}_t^2$ 和 $\boldsymbol{X}_{t-1}$ 之间的多重共线性(multicollinearity). 特别地，式(4.41)的线性回归分成两步. 在 2(a)步，移除 $\hat{x}_t^2$ 对 $\boldsymbol{X}_{t-1}$ 的线性依赖关系通过拟合回归

$$\hat{x}_t^2=\boldsymbol{X}'_{t-1}\boldsymbol{\beta}+u_t$$

得到残差 $\hat{u}_t=\hat{x}_t^2-\boldsymbol{X}_{t-1}\hat{\boldsymbol{\beta}}$. 在 2(b)步，考虑线性回归

$$\hat{a}_t=\hat{u}_t\boldsymbol{\alpha}+v_t$$

156

得到平方残差和 $SSR_1=\sum_{t=p+1}^{T}(\hat{a}_t-\hat{u}_t\hat{\alpha})^2=\sum_{t=p+1}^{T}\hat{v}_t^2$ 来检验原假设 $\alpha=0$.

4.2.2.2 F 检验

为了改进 Keennan 与 RESET 检验的功效，Tsay (1986)选择了不同的回归量 M_{t-1}. 特别地，他建议使用 $\boldsymbol{M}_{t-1}=\text{vech}(\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1})$，其中 vech$(A)$表示矩阵 A 的半拉直向量，即 $\boldsymbol{A}$ 对角线及其下面的元素组成的向量，见第 8 章附录 B 的更多内容. 例如，当 $p=2$ 时，$\boldsymbol{M}_{t-1}=(x_{t-1}^2, x_{t-1}x_{t-2}, x_{t-2}^2)'$. 对 AR($p$)模型 $\boldsymbol{M}_{t-1}$ 的维数是 $p(p+1)/2$. 在实际中，这个检验法就是对如下线性最小二乘回归用普通偏 F 统计量来检验 $\boldsymbol{\alpha}=0$：

$$x_t=\boldsymbol{X}'_{t-1}\boldsymbol{\phi}+\boldsymbol{M}'_{t-1}\boldsymbol{\alpha}+e_t$$

其中 e_t 是误差项. 在 x_t 是线性 AR(p)过程的假定下，偏 F 统计量服从一个自由度为 g 和 $T-p-g-1$ 的 F 分布，其中 $g=p(p+1)/2$. 我们把这种 F 检验称为 Ori-F 检验. Luukkonen, Saikkonen, 和 Teräsvirta(1988)进一步发展了这个检验法，将 $\boldsymbol{M}_{t-1}$ 增加到了有三次项 x_{t-i}^3(对 $i=1, \cdots, p$).

4.2.2.3 门限检验

在研究中，当备择模型是 SETAR 模型时，可以选特别的检验统计量来增加检验的功效. 这些特别的检验之一就是似然比统计量. 不过，这个检验遇到了在线性的原假设下没有定义参数的困难，这是因为对线性 AR 过程，门限是没有定义的. 另一种特殊的检验试图将检验门限非线性转化成探测模型的变化. 对门限非线性讨论这两种检验的区别是有趣的.

为简化讨论起见，考虑一个简单的例子. 备择模型是门限变量为 x_{t-d} 的 2 体制 SETAR 模型. 原假设是 H_0：x_t 服从线性 AR(p)模型

$$x_t=\phi_0+\sum_{i=1}^{p}\phi_i x_{t-i}+a_t \tag{4.43}$$

而备择假设是 $H_a: x_t$ 服从 SETAR 模型

$$x_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p} \phi_i^{(1)} x_{t-i} + a_{1t} & \text{如果 } x_{t-d} < r_1 \\ \phi_0^{(2)} + \sum_{i=1}^{p} \phi_i^{(2)} x_{t-i} + a_{2t} & \text{如果 } x_{t-d} \geqslant r_1 \end{cases} \tag{4.44}$$

其中 r_1 是门限. 对给定的实现 $\{x_t\}_{t=1}^{T}$ 且假定是正态的, 令 $l_0(\hat{\phi}, \hat{\sigma}_a^2)$ 为对数似然函数, 它是以 $\phi=(\phi_0, \cdots, \phi_p)'$ 和 σ_a^2 的最大似然估计计算得到的. 这很容易计算. 如果门限 r_1 给定, 那么在备择条件下似然函数也是易于计算的. 令 $l_1(r_1; \hat{\phi}_1, \hat{\sigma}_1^2; \hat{\phi}_2, \hat{\sigma}_2^2)$ 为对数似然函数, 它是在知道门限 r_1 的条件下, 由 $\phi_i=(\phi_0^{(i)}, \cdots, \phi_p^{(i)})'$ 和 σ_i^2 的最大似然估计计算得到的. 对数似然比定义为

[157]

$$l(r_1) = l_1(r_1; \hat{\phi}_1, \hat{\sigma}_1^2; \hat{\phi}_2, \hat{\sigma}_2^2) - l_0(\hat{\phi}, \hat{\sigma}_a^2)$$

它是未知门限 r_1 的函数. 然而在原假设下, 并没有门限, 而且 r_1 没有定义. 在原假设下, 参数 r_1 称为讨厌参数(nuisance parameter). 因此, 似然比的渐近分布与通常的似然比统计量有很大不同. 见 Chan (1991)以获得更多的细节和这个检验的临界值. 通常的做法是取 $l_{\max} = \sup_{v<r_1<u} l(r_1)$ 作为检验统计量, 其中 v 和 u 是事先指定为门限的下界和上界. Davis (1987); Andrews 和 Ploberger (1994) 进一步讨论了原假设下含讨厌参数的假设检验. 模拟法经常用于获取检验统计量 $l_{\max}$ 的经验临界值, 而 $l_{\max}$ 依赖于 v 和 u 的选择. $l(r_1)$ 在 $r_1 \in [v, u]$ 上的平均值也被 Andrews 和 Ploberger 考虑作为检验统计量.

Tsay (1989)利用排序自回归(arranged autoregression)和递归估计得到一个替换门限非线性的检验. 在备择假设 H_a 下, 排序自回归寻求将 SETAR 模型问题转变成一个模型改变问题, 其改变点就是门限 r_1. 为说明这点, 式(4.44)的 SETAR 模型说明 x_t 实质上服从两个线性模型, 依赖 $x_{t-d}<r_1$ 或是 $x_{t-d} \geqslant r_1$. 对实现值 $\{x_t\}_{t=1}^{T}$, x_{t-d} 取值 $\{x_1, \cdots, x_{T-d}\}$. 令 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(T-d)}$ 为 $\{x_t\}_{t=1}^{T-d}$ 排列好的数据(比如将这些观测值按递增的顺序排好). 这样 SETAR 模型可写成:

$$x_{(j)+d} = \beta_0 + \sum_{i=1}^{p} \beta_i x_{(j)+d-i} + a_{(j)+d}, \qquad j = 1, \cdots, T-d \tag{4.45}$$

其中 $\beta_i = \phi_i^{(1)}$ 当 $x_{(j)} < r_1$ 时; $\beta_i = \phi_i^{(2)}$ 当 $x_{(j)} \geqslant r_1$ 时. 从而, 门限 r_1 是式(4.45)的线性回归的改变点, 称式 (4.45)为排序自回归(门限 x_{t-d} 是递增顺序的). 注意到式(4.45)的排序自回归并没改变 x_t 对 x_{t-i}(对 $i=1, \cdots, p$)的动态依赖性, 这是因为 $x_{(j)+d}$ 仍然依赖于 $x_{(j)+d-i}$(对 $i=1, \cdots, p$). 上面完成了将 SETAR 模型放到门限空间中以代替时间空间. 这就是说, 带有较小的 x_{t-d} 的式子会出现在带有较大 x_{t-d} 的式子之前. Tsay (1989)的门限检验如下:

第 1 步: 用 $j=1, \cdots, m$ 拟合式(4.45), 其中 m 是事先指定的正整数(比如 30). 记 β_i 的最小二乘估计是 $\hat{\beta}_{i,m}$, 其中 m 表示用于估计的数据点个数.

[158]

第 2 步: 计算预报残差

$$\hat{a}_{(m+1)+d} = x_{(m+1)+d} - \hat{\beta}_{0,m} - \sum_{i=1}^{p} \hat{\beta}_{i,m} x_{(m+1)+d-i}$$

和它的标准差. 令$\hat{e}_{(m+1)+d}$为标准化的预报残差.

第 3 步：用递归最小二乘法去修正最小二乘估计$\hat{\beta}_{i,m+1}$，通过合并新数据点 $x_{(m+1)+d}$实现.

第 4 步：重复第 2、3 步，直至所有数据点都处理过.

第 5 步：考虑标准化的预报残差的线性回归

$$\hat{e}_{(m+j)+d} = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{(m+j)+d-i} + v_t, \qquad j = 1,\cdots,T-d-m \tag{4.46}$$

计算通常的 F 统计量以检验式(4.46) 中的 $\alpha_i=0(i=0, \cdots, p)$. 在 x_t 服从线性 AR(p)模型的原假设下，F 比有一个自由度为 $p+1$ 和 $T-d-m-p$ 的 F 分布. 我们称上面的 F 检验为 Tar-F 检验. 这个检验背后的思想是在原假设下式(4.45)的排序自回归中没有模型改变，所以标准化的预测残差应该接近零均值、方差为 1 的独立同分布的随机变量. 在这种情况下，那些标准化的预测残差应该与回归值 $x_{(m+j)+d-i}$不相关. 要了解更多关于递归最小二乘法和 Tar-F 检验表现的模拟研究的细节，请参见 Tsay (1989). Tar-F 检验回避了似然比检验中遇到的讨厌参数问题，它不要求知道门限 r_1. 它只是简单地检验了在原假设下，预报残差与回归值的不相关性. 因此，这个检验不依赖于一定要知道备择模型中体制的数目. 但是如果真实模型确实是有已知更新分布的 2 体制 SETAR 模型，那么 Tar-F 检验的功效不如似然比检验.

4.2.3 应用

在这一节中，我们把先前讨论的一些非线性检验应用到五个时间序列中. 对一个真实的金融时间序列，用 AR 模型移除数据中的相关成分，再将检验用到模型的残差序列中. 这五个序列如下：

1)r_{1t}：500 个观测值组成的独立同分布的、服从 $N(0, 1)$的模拟序列.

2)r_{2t}：独立同分布的、服从自由度为 6 的学生-t 分布的模拟序列，样本容量是 500. [159]

3)a_{3t}：1926 年到 1997 年的 CRSP 相等加权指数的月对数收益率的 864 个观测值的残差序列. 用到的线性 AR 模型是

$$(1-0.180B+0.099B^3-0.105B^9)r_{3t} = 0.0086 + a_{3t}$$

4)a_{4t}：1926 年到 1997 年的 CRSP 的价值加权指数的月对数收益率的 864 个观测值的残差序列. 用到的线性 AR 模型是

$$(1-0.098B+0.111B^3-0.088B^5)r_{4t} = 0.0078 + a_{4t}$$

5)a_{5t}：1926 年到 1997 年的 IBM 股票月对数收益率的 864 个观测值的残差序列. 用到的线性 AR 模型是

$$(1-0.077B)r_{5t} = 0.011 + a_{5t}$$

表 4-2 是非线性检验的结果. 对模拟序列和 IBM 收益率，F 检验是建立在 AR(6)模型的基础上. 对指数收益率，AR 的阶与前面的模型一样. 对 BDS 检验，我们选择 $\delta=\hat{\sigma}_a$ 和 $\delta=1.5\hat{\sigma}_a$，k 取 2，…，5. 表中还给出了 Ljung-Box 统计量以确定在应用非线性检验前残差序列没有序列

相关性. 与它们的渐近临界值比较，BDS 检验和 F 检验对模拟序列在 5%的水平下是不显著的，不过对真实的金融时间序列而言，BDS 检验却是高度显著的. F 检验对指数收益率也显示出了显著的结果，但对 IBM 对数收益率的非线性没有检验出来. 总之，这些检验证明了模拟序列是线性的，并认为股票收益率是非线性的.

表 4-2

Data	Q	Q	BDS($\delta=1.5\hat{\sigma}_a$)			
	(5)	(10)	2	3	4	5
N(0, 1)	3.2	6.5	−0.32	−0.14	−0.15	−0.33
t_6	0.9	1.7	−0.87	−1.18	−1.56	−1.71
ln(ew)	2.9	4.9	9.94	11.72	12.83	13.65
ln(vw)	1.0	9.8	8.61	9.88	10.70	11.29
ln(ibm)	0.6	7.1	4.96	6.09	6.68	6.82
Data		$d=1$	BDS($\delta=\hat{\sigma}_a$)			
	Ori-F	Tar-F	2	3	4	5
N(0, 1)	1.13	0.87	−0.77	−0.71	−1.04	−1.27
t_6	0.69	0.81	−0.35	−0.76	−1.25	−1.49
ln(ew)	5.05	6.77	10.01	11.85	13.14	14.45
ln(vw)	4.95	6.85	7.01	7.83	8.64	9.53
ln(ibm)	1.32	1.51	3.82	4.70	5.45	5.72

注：对模拟序列和一些股票对数收益率的非线性检验. 模拟序列的样本容量是 500，股票收益率的样本容量是 864. BDS 检验用 $k=2, \cdots, 5$.

160

4.3 建模

非线性时间序列的建模必须包含一些主观判断. 不过，仍然要遵守一些总体上的准则. 非线性检验开始时要建立在一个合适的线性模型的基础上. 对金融时间序列，Ljung-Box 统计量和 Engle 检验通常用于检测条件异方差性. 对一般序列，4.2 节的其他检验法可以用. 如果非线性在统计上是显著的，那么可选择接受一类非线性模型. 这种选择可能依赖分析者的经验和研究的问题的实质. 对于波动率模型，ARCH 过程的阶是通过检查平方序列的偏自相关函数确定的. 对 GARCH 和 EGARCH 模型，在绝大多数应用中，只考虑如同(1, 1)、(1, 2)和(2, 1)低阶的情况. 更高阶的模型很难估计和理解. 对 TAR 模型，可以用 Tong (1990)和 Tsay(1989, 1998) 给出的步骤程序建立合适的模型. 当样本容量充分大时，可以用非参数技术去探测数据的非线性因素，然后据此选择合适的非线性模型；见 Chen 和 Tsay (1993)，Cai、Fan 和 Yao(1999). Lewis 和 Stevens(1991)的 MARS 程序也可以用来探测数据的动态结构. 最后，像 Akaike 信息准则 (Akaike, 1974) 这样的信息标准以及广义机会比(odd ratios) (Chen, McCulloch 和 Tsay (1997))可以用来区分候选的非线性模型. 选定的模型在被用来预测前应该仔细检查核对.

4.4 预测

与线性模型不同，当预测长度大于 1 时，对绝大多数非线性模型计算预测值并没有显示表达公式．我们用参数自助(parameric bootstraps)法计算非线性预测值．我们容易理解模型在用于预测前需要严格的检查，以判断对所研究的序列是否合适．通过一个模型，我们可以了解动态结构和新息分布．在一些情况下，我们可以把估计的参数看作已知的．

4.4.1 参数自助法

令 T 为预测原点，l 为预测步长($l>0$)．这就是说，我们处在时间指标 T 处并对预测 x_{T+l} 感兴趣．参数自助法计算实现值 x_{T+1}，…，x_{T+l}依次通过下面步骤：(a)由模型指定的新息分布得到一个新的新息．(b)用模型、数据以及以前的预测 x_{T+1}，…，x_{T+i-1} 计算 x_{T+i}．这样得到 x_{T+l}的一个实现值．重复这个过程 M 次，得到 x_{T+l}的 M 个实现值，用$\{x_{T+l}^{(j)}\}_{j=1}^{M}$表示．$x_{T+l}$的预测点取 $x_{T+l}^{(j)}$的样本均值．令预测为 $x_T(l)$．在一些应用中我们取 $M=3\,000$，效果不错．实现值$\{x_{T+l}^{(j)}\}_{j=1}^{M}$也可用于获得 x_{T+l}的经验分布．我们稍后利用经验分布去评估预测功效． 161

4.4.2 预测的评估

评估模型的预测功效有很多方法．从方向度量到大小度量再到分布度量．方向度量考虑将来的变动方向(上升或下降)由模型可推知．预测明天的标准普尔 500 指数将要上升或下降是有实际意义的方向预测的例子．预测年终日标准普尔 500 指数的价值属于大小度量的例子．最后，评估从现在到年终的日标准普尔 500 指数将上升 10%或更多的可能性，要求知道将来该指数的条件概率分布．评估这样一个估价的精度则需要分布度量．

在实际中，可利用的数据集被分成两部分子样本．第一部分子样本的数据用来建立一个非线性模型，第二部分子样本用来评估模型的预测功效．我们称这两部分子样本分别为估计和预测子样本．在一些研究中，滚动预测程序在使用着，当预测原点前进时，一个新的数据点从预测子样本转移到估计子样本中．在下文中，我们简要地讨论文献中一些通常用到的预测功效的度量方法．不过要记住在比较模型优劣时，不存在单独的广泛接受的度量方法．为了更好了解这种比较，可能需要一个基于预测目的的效用函数．

4.4.2.1 方向度量

一个典型的度量法是用一个 2×2 的列联表(contingency table)来总结在预测子样本中预测 x_{T+l}的上升和下降相对应的模型的“击中”(hits)和“丢失”(misses)数目．具体说来，列联表如右给出：

实际	预测		
	上升	下降	
上升	m_{11}	m_{12}	m_{10}
下降	m_{21}	m_{22}	m_{20}
	m_{01}	m_{02}	m

其中 m 是预测子样本中 l 步向前预测的总数目．m_{11}是预测向上运动且“击中”(预测正确——译者注)的数目，m_{21}是预测市场向下运动且“丢失”(预测错误——译者注)的数目，依此类推． 162

m_{11} 和 m_{22} 有较大值，意味着较好的预测. 其检验统计量

$$\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(m_{ij}-\frac{m_{i0}m_{0j}}{m}\right)^2}{\frac{m_{i0}m_{0j}}{m}}$$

可以用来评估这个模型的功效. 大的 χ^2 值表示该模型优于随机选择的机会. 在一些适度的条件下，χ^2 渐近服从自由度为 1 的 χ^2 分布. 有关这种度量的更进一步讨论见 Dahl 和 Hylleberg (1999).

作为方向度量的例子，考虑图 4-6 中的 8-4-1 前馈神经网络的 1 步向前概率预测. 这个 2×2 的表示网络“击中”和“丢失”的表格是右图. 这个表显示了该网络对向上运动的预测不错，但对股票向下运动的预测是不理想的. 这个表格的 χ^2 统计量是 0.137，p 值是 0.71. 从而，网络并没有显著地优于一个对“向上”“向下”运动等概率的随机游动模型.

实际	预测		
	上升	下降	
上升	12	2	14
下降	8	2	10
	20	4	24

4.4.2.2 大小度量

有三个统计量经常用于度量点预测的功效. 它们是均方误差(MSE)，平均绝对偏差(MAD)和平均绝对百分比误差(MAPE). 对 l 步向前预测，这些度量定义如下

$$MSE(l)=\frac{1}{m}\sum_{j=0}^{m-1}[x_{T+l+j}-x_{T+j}(l)]^2 \tag{4.47}$$

$$MAD(l)=\frac{1}{m}\sum_{j=0}^{m-1}|x_{T+l+j}-x_{T+j}(l)| \tag{4.48}$$

$$MAPE(l)=\frac{1}{m}\sum_{j=0}^{m-1}\left|\frac{x_{T+j}(l)}{x_{T+j+l}}-1\right| \tag{4.49}$$

[163] 其中 m 是预测子样本中可用于 l 步向前预测的数目. 在应用中，通常选择上面三种度量之一，而在这种度量下最小的模型认为是最好的 l 步向前预测模型. l 的不同可能会导致选用不同的模型. 这些度量在模型对比中仍然有其他的局限性，例如，见 Clements 和 Hendry(1993).

4.4.2.3 分布度量

从业人员最近开始用预测分布对一个模型的预测功效进行评估. 严格地讲，预测分布融合了在预测时参数的不确定性. 如果参数看作固定，称为条件预测分布. 通过参数自助法得到的 x_{T+l} 的经验分布是一个条件预测分布. 这个经验分布通常用于计算分布度量. 令 $u_T(l)$ 为观测值 x_{T+l} 在前面所述经验分布中的百分比位置点. 我们有一个 m 个百分比位置点 $\{u_{T+j}(l)\}_{j=0}^{m-1}$ 的集合，其中 m 是预测子样本中 l 步向前预测的数目. 如果接受的模型合适，$\{u_{T+j}(l)\}$ 应该是[0, 1]上均匀分布的一个随机样本. 对一个充分大的 m，可以计算[0, 1]上均匀分布的 $\{u_{T+j}(l)\}$ 的科尔莫戈罗夫-Smirnov 统计量. 这个统计量可用于模型的检验和预测的比较.

4.5 应用

在这一小节中，我们通过分析 1948 年到 1993 年经季节调整后的每季度美国平民失业率来说明非线性时间序列模型. 这个序列 Montgomery、Zarnowitz、Tsay 和 Tiao (1998)详细地分析过，这里我们用非线性模型重复他们的一些分析内容. 图 4-8 是数据的时间图. 这个序列明显的特征包括：(a)与美国商业周期相反周期性，(b)失业率上升很快，下降却很慢. 这后一特征说明该序列的动态结构是非线性的.

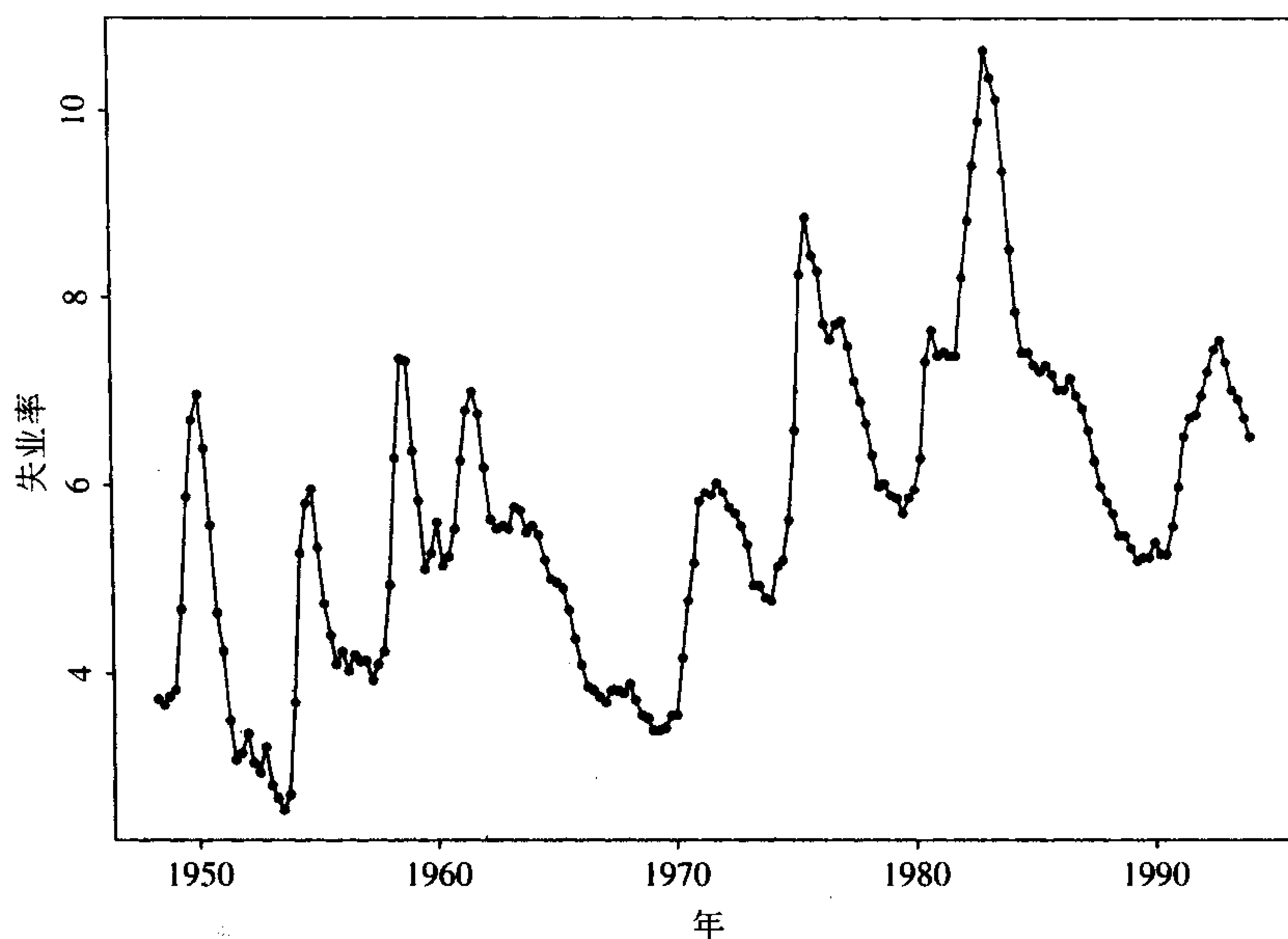


图 4-8 从 1948 年到 1993 年季度美国失业率(经季度调整后)的时间图

记该序列为 x_t，并且令 $\Delta x_t = x_t - x_{t-1}$ 为失业率的变动值. Montgomery 等(1998)建立了线性模型

$$(1-0.31B^4)(1-0.65B)\Delta x_t = (1-0.78B^4)a_t, \quad \hat{\sigma}_a^2 = 0.090 \tag{4.50}$$

其中三个系数的标准差分别为 0.11，0.06 和 0.07. 虽然数据经过了季节调整，但模型还是呈现了季节性. 这说明季节调整并没有成功的移除季节性. 我们把这个模型当作预测比较的基准模型.

为检验非线性，我们对由差分序列 Δx_t 建立的 AR(5)模型进行 4.2 节中介绍过的一些非线性检验. 结果在表 4-3 给出. 所有的检验法都拒绝了线性假设. 实际上，所有用 AR(p)(其中 $p=2, \cdots, 10$)模型的线性假设都被拒绝.

表 4-3

Type	Ori-F	LST	TAR(1)	TAR(2)	TAR(3)	TAR(4)
Test	2.80	2.83	2.41	2.16	2.84	2.98
p value	0.0007	0.0002	0.0298	0.0500	0.0121	0.0088

注：1948 年第二季度到 1993 年第四季度的美国季度失业率的变化的非线性检验. 使用 AR(5)模型，LST 表示 Luukkonen 等(1988)提出的检验，TAR(d)指的是延迟为 d 的门限检验.

用类似于 Tsay (1989)的建模程序，Montgomery 等(1998)对 Δx_t 序列建立了下面的 TAR 模型：

$$\Delta x_t = \begin{cases} 0.01 + 0.73\Delta x_{t-1} + 0.10\Delta x_{t-2} + a_{1t}, & \text{如果 } \Delta x_{t-2} \leqslant 0.1 \\ 0.18 + 0.80\Delta x_{t-1} - 0.56\Delta x_{t-2} + a_{2t}, & \text{否则} \end{cases} \quad (4.51)$$

a_{1t}和 a_{2t}的样本方差分别是 0.76 和 0.165，体制 1 的三个系数的标准差分别是 0.03、0.10 和 0.12，体制 2 的三个系数的标准差则分别是 0.09、0.1 和 0.16. 这个模型说明了美国季度失业率的变化值 Δx_t 在 $x_{t-2}-x_{t-3}$、门限为 0.1 的基准空间中，很像一个分段线性模型. 直观上，该模型暗示了随着最近失业率的改变，失业人数的表现是不同的. 在第一个体制下，失业率下降或者有轻微的上升. 这样经济应该是稳定的，实质上变动率服从一个简单的 AR(1)模型，因为 2 阶延迟系数是不显著的. 在第二个体制下，失业率(0.1 或更大)有一个大幅跳跃，这典型地与经济周期的紧缩期相对应. 而且这个时期可能发生政府干预经济以及产业结构调整. 这里 Δx_t 服从一个带正常数项的 AR(2)模型，显示 x_t 向上的趋势. 这个 AR(2)多项式有两个复特征根，显示了 Δx_t 可能的周期行为. 从而，x_t 出现转折点的可能性增大，意味着 x_t 大幅增长的时期会很短. 这意味着美国经济的紧缩期比扩张期要短些.

164~165

Montgomery 等(1998) 用马尔可夫链蒙特卡罗方法得到下面关于 Δx_t 的马尔可夫转换模型：

$$\Delta x_t = \begin{cases} -0.07 + 0.38\Delta x_{t-1} - 0.05\Delta x_{t-2} + \varepsilon_{1t}, & \text{如果 } s_t = 1 \\ 0.16 + 0.86\Delta x_{t-1} - 0.38\Delta x_{t-2} + \varepsilon_{2t}, & \text{如果 } s_t = 2 \end{cases} \quad (4.52)$$

Δx_t 的条件均值当 $s_t=1$ 时是 -0.10，当 $s_t=2$ 时是 0.31. 从而，第一个状态表示经济的扩张期，第二个状态表示经济的紧缩期. ε_{1t}和 ε_{2t}的样本方差分别是 0.031 和 0.192. 在状态 $s_t=1$ 时，三个参数的标准差分别是 0.03、0.14 和 0.11，而在状态 $s_t=2$ 时，它们则分别是 0.04、0.13 和 0.14. 状态转移概率是 $P(s_t=2 \mid s_{t-1}=1)=0.084(0.060)$ 和 $P(s_t=1 \mid s_{t-1}=2)=0.126(0.053)$，其中括号里的数字是对应的标准差. 这个模型说明在第二个状态下，当 AR(2)多项式有复特征根时，失业率 x_t 有向上的趋势. 模型的这个特征很像式(4.51)的第 2 个体制的 TAR 模型. 在第一个状态下，失业率 x_t 有弱得多的自回归结构的轻微的下降趋势.

预测功效

Montgomery 等(1998) 用一个滚动程序预测失业率 x_t. 该程序如下

1)以 $T=83$ 为预测原点，对应于 1968 年第二季度，它在文献中用于监视各种预测失业率的经济度量模型的功效. 用 1948 年第一季度到预测原点(包括预测原点)的数据估计线性、TAR 和 MSA 模型.

2)进行1季度和5季度向前预测，并计算每个模型的预测误差．对非线性模型的预测使用4.4节中的参数自助法计算．

3)预测原点向前移1步，然后重复估计和预测过程，直到每个数据都使用过．

4)用MSE和平均预测误差来比较模型的功效．

166

从表4-4可以看到式(4.50)的线性模型、式(4.51)的TAR模型和式(4.52)的MSA模型的预测相对MSE和平均预测误差．用线性模型作为基准．比较是基于总体表现和在预测原点美国经济状况．从这张表格，我们观察到下面的结论：

表 4-4

(A) 模型	预测的相关 MSE				
	1步	2步	3步	4步	5步
	(a)总体比较				
线性	1.00	1.00	1.00	1.00	1.00
TAR	1.00	1.04	0.99	0.98	1.03
MSA	1.19	1.39	1.40	1.45	1.61
MSE	0.08	0.31	0.67	1.13	1.54
	(b)经济紧缩的预测原点				
线性	1.00	1.00	1.00	1.00	1.00
TAR	0.85	0.91	0.83	0.72	0.72
MSA	0.97	1.03	0.96	0.86	1.02
MSE	0.22	0.97	2.14	3.38	3.46
	(c)经济扩张的预测原点				
线性	1.00	1.00	1.00	1.00	1.00
TAR	1.06	1.13	1.10	1.15	1.17
MSA	1.31	1.64	1.73	1.84	1.87
MSE	0.06	0.21	0.45	0.78	1.24
(B) 模型	**预测误差的均值**				
	1步	2步	3步	4步	5步
	(a)总体比较				
线性	0.03	0.09	0.17	0.25	0.33
TAR	−0.10	−0.02	−0.03	−0.03	−0.01
MSA	0.00	−0.02	−0.04	−0.07	−0.12
	(b)经济紧缩的预测原点				
线性	0.31	0.68	1.08	1.41	1.38
TAR	0.24	0.56	0.87	1.01	0.86
MSA	0.20	0.41	0.57	0.52	0.14
	(c)经济扩张的预测原点				
线性	−0.01	0.00	0.03	0.08	0.17
TAR	−0.05	−0.11	−0.17	−0.19	−0.14
MSA	−0.03	−0.08	−0.13	−0.17	−0.16

注：对美国季度失业率用线性、MSA和TAR模型的样本外预测的比较．预测原点从1968年第2季度开始．用“MSE”标记的行表示基准线性模型的MSE(均方误差)．

167

1)总体上比较，TAR 模型和线性模型的 MSE 非常接近，但 TAR 模型的偏差更小. 而 MSA 模型有最大的 MSE，最小的偏差.

2)当预测原点在经济紧缩期时，TAR 模型无论是 MSE 还是偏差都比线性模型表现得好. MSA 模型也比线性模型好，但没有 TAR 模型好的那么多.

3)当预测原点在经济扩张期时，线性模型比两个非线性模型要好.

结果表明当美国经济处于紧缩期时，在预测美国季度失业率上，非线性模型的表现超过线性模型. 这并不奇怪，就像前面提到的，在经济紧缩时，政府会干预经济，并且会进行产业结构调整. 这些外部事件引起了美国失业率的非线性. 直观上，这样的改进很重要，因为在经济紧缩期，人们会更关注经济预测.

附录 A　一些关于非线性波动率模型的 RATS 程序

A. 用于给 IBM 股票日对数收益率评估 AR(2)-TAR-GARCH(1, 1)模型，数据文件是“d-ibmln99. dat”

```
all 0 9442:1
open data d-ibmln99.dat
data(org=obs) / rt
set h = 0.0
*nonlin mu p1 p2 a0 a1 a2 b0 b1 b2
nonlin mu p2 a1 a2 b0 b1 b2
*frml at = rt(t)-mu-p1*rt(t-1)-p2*rt(t-2)
frml at = rt(t)-mu-p2*rt(t-2)
frml u = (at(t-1)/abs(at(t-1))+1.0)/2.0
frml gvar1 = a1*at(t-1)**2+a2*h(t-1)
frml gvar = gvar1(t)+u(t)*(b0+b1*at(t-1)**2+b2*h(t-1))
frml garchln = -0.5*log(h(t)=gvar(t))-0.5*at(t)**2/h(t)
smpl 4 9442
compute mu = 0.03, p1 = 0.1, p2 = -0.03
compute a0 = 0.1, a1 = 0.1, a2 = 0.6, b0 = 0.1, b1 = 0.05
compute b2 = 0.1, a3 = 0.1, b3 = 0.1
maximize(method=simplex,iterations=10) garchln
smpl 4 9442
maximize(method=bhhh,recursive,iterations=150) garchln
set fv = gvar(t)
set resid = at(t)/sqrt(fv(t))
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

168

B. 用于给 3M 股票月简单收益率评估一个平滑 TAR 模型，数据文件是“m-mmm. dat”

```
all 0 623:1
open data m-mmm.dat
data(org=obs) / mmm
set h = 0.0
nonlin a0 a1 a2 a00 a11 mu
frml at = mmm(t) - mu
frml var1 = a0+a1*at(t-1)**2+a2*at(t-2)**2
frml var2 = a00+a11*at(t-1)**2
frml gvar = var1(t)+var2(t)/(1.0+exp(-at(t-1)*1000.0))
frml garchlog = -0.5*log(h(t)=gvar(t))-0.5*at(t)**2/h(t)
smpl 3 623
compute a0 = .01, a1 = 0.2, a2 = 0.1
compute a00 = .01, a11 = -.2, mu = 0.02
maximize(method=bhhh,recursive,iterations=150) garchlog
set fv = gvar(t)
set resid = at(t)/sqrt(fv(t))
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

附录 B　神经网络的 S-Plus 命令

以下 S-Plus 命令是用来建立例 4.5 中的 3-2-1 跳跃层的前馈网络. 带“#”号的行是注释. 数据文件是“m-ibmln. dat”.

```
# load the data into S-Plus workspace.
x_scan(file='m-ibmln.dat')
# select the output: r(t)
y_x[4:864]
# obtain the input variables: r(t-1), r(t-2), and r(t-3)
ibm.x_cbind(x[3:863],x[2:862],x[1:861])
# build a 3-2-1 network with skip layer connections
# and linear output.
ibm.nn_nnet(ibm.x,y,size=2,linout=T,skip=T,maxit=10000,
decay=1e-2,reltol=1e-7,abstol=1e-7,range=1.0)
# print the summary results of the network
summary(ibm.nn)
# compute \& print the residual sum of squares.
sse_sum((y-predict(ibm.nn,ibm.x))^2)
print(sse)
#eigen(nnet.Hess(ibm.nn,ibm.x,y),T)$values
# setup the input variables in the forecasting subsample
ibm.p_cbind(x[864:887],x[863:886],x[862:885])
# compute the forecasts
yh_predict(ibm.nn,ibm.p)
# The observed returns in the forecasting subsample
yo_x[865:888]
# compute \& print the sum of squares of forecast errors
ssfe_sum((yo-yh)^2)
print(ssfe)
# quit S-Plus
q()
```

169

练习题

1. 考虑 General Electric(GE)股票从 1926 年 1 月到 1999 年 12 月的月对数收益率. 可以从 CRSP 下载数据或使用网上的文件"m-ge2699. dat". 文中的对数收益率是以百分比形式给出的. 设 a_{t-1}是 $t-1$ 时刻的抖动，给上述数据建立一个以 a_{t-1}为门限变量、0 为门限的门限 GARCH 模型，并检验所拟合的模型.
2. 假设 GE 股票的月对数收益率(以百分比形式)服从一个平滑的门限 GARCH(1, 1)模型. 对样本时间区间为 1926 年 1 月至 1999 年 12 月的数据，拟合的模型为

$$r_t = 1.06 + a_t, \quad a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = 0.103a_{t-1}^2 + 0.952\sigma_{t-1}^2 + \frac{1}{1+\exp(-10a_{t-1})}(4.490 - 0.193\sigma_{t-1}^2)$$

其中所有的估计值都是高度显著的，指数中的系数 10 是事先给定的以简化估计过程，$\{\varepsilon_t\}$是独立同分布、服从 $N(0, 1)$的序列. 假设 $a_{888}=16.0$ 和 $\sigma_{888}^2=50.2$，1 步向前波动率预测值$\hat{\sigma}_{888}^2(1)$是什么? 如果换成 $a_{888}=-16.0$，那么 1 步向前波动率预测值$\hat{\sigma}_{888}^2(1)$又是什么?

3. 假设一支股票的月对数收益率(以百分比形式)服从如下的马尔可夫转换模型

$$r_t = 1.25 + a_t, \quad a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \begin{cases} 0.10a_{t-1}^2 + 0.93\sigma_{t-1}^2, & \text{如果 } s_t = 1 \\ 4.24 + 0.10a_{t-1}^2 + 0.78\sigma_{t-1}^2, & \text{如果 } s_t = 2, \end{cases}$$

[170] 其中转移概率为

$$P(s_t = 2 \mid s_{t-1} = 1) = 0.15, \quad P(s_t = 1 \mid s_{t-1} = 2) = 0.05.$$

假设 $a_{100}=6.0$，$\sigma_{100}^2=50.0$ 和以概率 1.0 有 $s_{100}=2$，以 $t=100$ 为预测原点的 1 步向前波动率预测值是什么? 如果 $s_{100}=2$ 的概率减小到 0.8，那么，以 $t=100$ 为预测原点的 1 步向前波动率预测值又是什么?

4. 再考虑 GE 股票从 1926 年 1 月到 1999 年 12 月的月对数收益率. 用 1998 年和 1999 年的收益率来进行预测评价.
   - 给此收益率序列拟合一个 3-2-1 的前馈神经网络，并计算预测子样本中的 1 步向前预测的均方误差. 记下估计子样本中网络的偏差和权重.
   - 假定我们欲预测向前 1 个月股票运动的方向. 用输出节点的 Heaviside 函数给此收益率序列拟合一个 6-5-1 的前馈神经网络. 计算预测子样本中的 1 步向前预测，并与实际运动相比较.
5. 因为在利率期限结构中存在逆收益曲线，利率分布是非线性的. 为了验证这个事实，考虑美国的两种周利率：(a)1 年期固定期限国库券，(b)三年期固定期限国库券. 如第 2 章一样，记这两种利率分别为 r_{1t}和 r_{3t}. 数据的时间区间为 1962 年 1 月 5 日 1999 年 9 月 10 日，数据可从网上文件"wgs3yr. dat"和"wgs1yr. dat" 中得到.

- 用 $s_t = r_{3t} - r_{1t}$ 表示对数利率之差. $\{s_t\}$ 是线性的吗? 进行非线性检验，并在 5％的显著水平下得出结论.
- 用 $s_t^* = (r_{3t} - r_{3,t-1}) - (r_{1t} - r_{1,t-1}) = s_t - s_{t-1}$ 表示利率差的变化. $\{s_t^*\}$ 是线性的吗? 进行非线性检验，并在 5％的显著水平下得出结论.
- 对 $\{s_t\}$ 建立一个门限模型，并检验所拟合的模型.
- 对 $\{s_t^*\}$ 建立一个门限模型，并检验所拟合的模型.

参考文献

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Andrews, D. W. K., and Ploberger, W. (1994), "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383–1414. [171]

Brock, W., Dechert, W. D., and Scheinkman, J. (1987), "A test for independence based on the correlation dimension," Working paper, Department of Economics, University of Wisconsin, Madison.

Brock, W., Hsieh, D. A., and LeBaron, B. (1991), *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*, MIT Press: Cambridge.

Bryson, A. E., and Ho, Y. C. (1969), *Applied Optimal Control*. Blaisdell: New York.

Cai, Z., Fan, J., and Yao, Q. (1999), "Functional-coefficient regression models for nonlinear time series," working paper, University of North Carolina, Charlotte.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992), "A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling," *Journal of the American Statistical Association*, 87, 493–500.

Chan, K. S. (1991), "Percentage points of likelihood ratio tests for threshold autoregression," *Journal of the Royal Statistical Society, Series B*, 53, 691–696.

Chan, K. S. (1993), "Consistency and limiting distribution of the least squares estimator of a continuous autoregressive model," *The Annals of Statistics*, 21, 520–533.

Chan, K. S., and Tong H. (1986), "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–190.

Chan, K. S., and Tsay, R. S. (1998), "Limiting properties of the conditional least squares estimator of a continuous TAR model," *Biometrika*, 85, 413–426.

Chen, C., McCulloch, R. E., and Tsay, R. S. (1997), "A unified approach to estimating and modeling univariate linear and nonlinear time series," *Statistica Sinica*, 7, 451–472.

Chen, R., and Tsay, R. S. (1991), "On the Ergodicity of TAR(1) processes," *Annals of Applied Probability*, 1, 613–634.

Chen, R., and Tsay, R. S. (1993a), "Functional-coefficient autoregressive models," *Journal of the American Statistical Association*, 88, 298–308.

Chen, R., and Tsay, R. S. (1993b), "Nonlinear additive ARX models," *Journal of the American Statistical Association*, 88, 955–967.

Chen, R., Liu, J., and Tsay, R. S. (1995), "Additivity tests for nonlinear autoregressive models," *Biometrika* (1995), 82, 369–383.

Chen, T., and Chen, H. (1995), "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems," *IEEE Transactions on Neural Networks* 6, 911–917.

Cheng, B., and Titterington, D. M. (1994), "Neural networks: a review from a statistical perspective," *Statistical Science* 9, 2–54.

Clements, M. P., and Hendry, D. F. (1993), "On the limitations of comparing mean square forecast errors," *Journal of Forecasting*, 12, 617–637.

Dahl, C. M., and Hylleberg, S. (1999), "Specifying nonlinear econometric models by flexible regression models and relative forecast performance," working paper, Department of Economics, University of Aarhus, Denmark.

Davis, R. B. (1987), "Hypothesis testing when a nuisance parameter is present only under the alternative," *Biometrika*, 74, 33–43.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987–1007.

Epanechnikov, V. (1969), "Nonparametric estimates of a multivariate probability density," *Theory of Probability and its applications*, 14, 153–158.

172

Fan, J. (1993), "Local linear regression smoother and their minimax efficiencies," *The Annals of Statistics*, 21, 196–216.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Granger, C. W. J., and Andersen, A. P. (1978), *An Introduction to Bilinear Time Series Models*. Vandenhoek and Ruprecht: Gottingen.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hamilton, J. D. (1990), "Analysis of time series subject to changes in regime," *J. Econometrics*, 45, 39–70.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press: New Jersey.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press: New York.

Hansen, B. E. (1997), "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics*, 1, 119–131.

Hinich, M. (1982), "Testing for Gaussianity and linearity of a stationary time series," *Journal of Time Series Analysis*, 3, 169–176,

Hornik, K. (1993), "Some new results on neural network approximation," *Neural Networks* 6, 1069–1072.

Hornik, K., Stinchcombe, M., and White, H. (1989), "Multilayer feedforward networks are universal approximators," *Neural Networks* 2, 359–366.

Hsieh, D. A. (1989), "Testing for nonlinear dependence in daily foreign exchange rates," *Journal of Business*, 62, 339–368.

Keennan, D. M. (1985), "A Tukey non-additivity-type test for time series nonlinearity," *Biometrika*, 72, 39–44.

Kitagawa, G. (1998), "A self-organizing state space model," *Journal of the American Statistical Association*, 93, 1203–1215.

Lewis, P. A. W., and Stevens, J. G. (1991), "Nonlinear Modeling of Time Series Using Multivariate Adaptive Regression Spline (MARS)," *Journal of the American Statistical Association*, 86, 864–877.

Liu, J., and Brockwell,P. J. (1988), "On the general bilinear time-series model," *Journal of Applied Probability*, 25, 553–564.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988), "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

McCulloch, R. E., and Tsay, R. S. (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968–978.

McCulloch, R. E., and Tsay, R. S. (1994), "Statistical inference of macroeconomic time series via Markov switching models," *Journal of Time Series Analysis*, 15, 523–539.

McLeod, A. I., and Li, W. K. (1983), "Diagnostic checking ARMA time series models using squared-residual autocorrelations," *Journal of Time Series Analysis*, 4, 269–273.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998), "Forecasting the U.S. Unemployment Rate," *Journal of the American Statistical Association*, 93, 478–493.

Nadaraya, E. A. (1964), "On estimating regression," *Theory and Probability Application*, 10, 186–190.

Petruccelli, J., and Woolford, S. W. (1984), "A threshold AR(1) model," *Journal of Applied Probability*, 21, 270–286.

Potter, S. M. (1995), "A nonlinear approach to U.S. GNP," *Journal of Applied Econometrics*, 10, 109–125.

Priestley, M. B. (1980), "State-dependent models: a general approach to nonlinear time series analysis," *Journal of Time Series Analysis*, 1, 47–71.

Priestley, M. B. (1988), *Non-linear and Non-stationary Time Series Analysis*, Academic Press: London.

Rabemananjara, R., and Zakoian, J. M. (1993), "Threshold ARCH models and asymmetries in volatility," *Journal of Applied Econometrics*, 8, 31–49.

Ramsey, J. B. (1969), "Tests for specification errors in classical linear least squares regression analysis," *Journal of the Royal Statistical Society, Series B*, 31, 350–371.

Ripley, B. D. (1993), "Statistical aspects of neural networks," in *Networks and Chaos—Statistical and Probablistic Aspects*, eds O. E. Barndorff-Nielsen, J. L. Jensen, and W. S. Kendall, pp. 40–123. Chapman and Hall: London.

Subba Rao, T., and Gabr, M. M. (1984), *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Lecture Notes in Statistics, 24. Springer-Verlag: New York.

Teräsvirta, T. (1994), "Specification, estimation, and evaluation of smooth transition autoregressive models," *Journal of the American Statistical Association*, 89, 208–218.

Tiao, G. C., and Tsay, R. S. (1994), "Some Advances in Nonlinear and Adaptive Modeling in Time Series," *Journal of Forecasting*, 13, 109–131.

Tong, H. (1978), "On a threshold model," in *Pattern Recognition and Signal Processing*, ed. C.H. Chen, Sijhoff & Noordhoff: Amsterdam.

Tong, H. (1983), *Threshold Models in Nonlinear Time Series Analysis*, Lecture Notes in Statistics, Springer-Verlag: New York.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Oxford University Press: Oxford.

Tsay, R. S. (1986), "Nonlinearity tests for time series," *Biometrika*, 73, 461–466.

Tsay, R. S. (1989), "Testing and modeling threshold autoregressive processes," *Journal of the American Statistical Association*, 84, 231–240.

Tsay, R. S. (1998), "Testing and modeling multivariate threshold models," *Journal of the American Statistical Association*, 93, 1188–1202.

Venables, W. N., and Ripley, B. D. (1999), *Modern Applied Statistics with S-Plus*, 3rd ed. Springer-Verlag: New York.

Watson, G. S. (1964), "Smooth regression analysis," *Sabkhya, Series A*, 26, 359–372.

Zakoian, J. M. (1994), "Threshold heteroscedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.

173

174

第5章 高频数据分析与市场微观结构

高频数据是指在细小的时间间隔上抽取的观测值，金融中常指以日或更小的时间间隔抽取的观测值. 由于数据的获得与处理方法的发展，这些高频数据目前是可以得到的，并且由于其在市场微观结构实证研究方面的重要性而受到广泛的关注. 金融中最极端的高频数据是证券市场中记录每一笔交易或贸易的数据，这里的时间通常是以秒为单位测量的. 纽约股票交易所(NYSE)的交易行情 TAQ 数据库包含了整理过的从 1992 年至今的所有证券交易的记录，它包括了在 NYSE、AMEX、NASDAQ 以及美国在纽约以外地区的证券交易所(regional exchange)的交易数据. 伯克利(Berkeley)期权数据库对于期权交易提供了从 1976 年 8 月至 1996 年 12 月的类似数据. 对于国内外其他许多证券以及市场的交易数据，都是连续收集和处理的. Wood(2000) 对于高频金融研究提供了一些历史观察法.

高频金融数据在研究与交易过程和市场微观结构相关的大量问题中都是很重要的. 可以用来比较不同交易系统(如 NYSE 的公开叫价系统和 NASDAQ 的计算机交易系统)在价格发现(price discovery)方面的有效性；还可以用来研究某只特定股票买卖报价的动态性(如 Hasbrouck，1999；Zhang，Russell，和 Tsay，2001b). 在一个指令驱动的股票市场(如台湾股市交易所)中，高频数据还可以用来研究指令动态，更有趣的是可以用来研究"是谁提供了市场的流动性"这样的问题. Cho，Russell，Tiao 和 Tsay(2000)利用在台湾股市交易所中交易的 340 多只股票在一天中的每 5 分钟的收益率，研究设定日股价上下限的影响，发现了向股价上限趋近磁效应的显著证据.

然而，高频数据还有一些低频数据中不会出现的独特特征，从而，对于这些数据的分析给金融经济学家与统计学家提出了新的挑战. 本章中，我们主要研究这些特殊的特征，考虑分析高频数据的方法，并且讨论所得结果的应用. 特别地，我们讨论非同步交易、买卖报价价差、持续期模型、大量小间隔上的价格运动、以及价格变化和与价格变化相关的交易之间的时间持续期的二元模型. 讨论的这些模型在其他的科学领域(如通讯和环境)研究中也是适用的.

175

5.1 非同步交易

我们以对非同步交易的讨论开始. 股票交易(如同 NYSE 中的一样)并不是同步发生的，不同的股票有着不同的交易频率；即使是同一种股票，其交易强度也是一小时一小时地、一天一天地变化的. 然而我们经常是对一个固定的时间间隔如一天、一周或者一个月来分析收益率序列. 对于日序列，股价指的是其收盘价格，即该股票在一个交易日内最后一次交易的价格，而股票最后一次交易的实际时间也是一天天变化的. 这样，如果我们假定日收益率序列在

24 小时里是等间隔的往往是不正确的. 实践证明，即使是在真实的收益率序列是前后独立的时候，这种假定可以导致股票收益率可预测性的错误的结论.

对于股票日收益率，非同步交易可以导致：(a)股票收益率之间的 1 步延迟交叉相关，(b)组合收益率的 1 步延迟序列相关，(c)某些情形下，单只股票收益率序列的负序列相关. 考虑股票 A 与 B. 假定这两只股票是独立的，并且股票 A 比股票 B 的交易频繁. 对于在某一天接近收盘时刻出现的一个特定的影响市场的消息，股票 A 比股票 B 更可能在同一天中显示出这个消息的效应，缘于 A 的交易更频繁. 该消息对股票 B 的效应最终也会出现，但是可能会被延迟到下一个交易日. 如果这种情况发生，则好像是股票 A 的收益率引导着股票 B 的收益率. 因此，尽管这两只股票是独立的，但是它们的收益率序列可能会显示出显著的 1 步延迟交叉相关性. 对于一个持有股票 A 与 B 的组合，前面的交叉相关将会变成一个显著的 1 步延迟序列相关.

更复杂的是，非同步交易也能引起单只股票错误的负序列相关. 可以利用文献中的一些模型来研究这种现象；参见 Campbell，Lo 和 MacKinlay(1997) 及其参考文献. 这里我们采用 Lo 和 MacKinlay(1990) 提出的模型的一个简化形式. 令 r_t 表示证券在 t 时刻的连续复合收益率. 为了简便，假定 $\{r_t\}$ 是一个独立同分布的随机变量序列，均值 $E(r_t)=\mu$、方差 $\text{Var}(r_t)=\sigma^2$. 对每个时间段，证券不交易的概率为 π，它是不随时间变化的，并且与 r_t 独立. 令 r_t^o 表示观测到的收益率，如果 t 时刻没有交易，则 $r_t^o=0$，因为此时没有可以利用的信息. 如果 t 时刻有一个交易，则我们定义 r_t^o 为从前一个交易开始的累积收益率(也就是，$r_t^o=r_t+r_{t-1}+\cdots+r_{t-k_t}$，其中 k_t 是满足在 $t-k_t$，$t-k_t+1$，…，$t-1$ 时刻没有交易发生的最大的非负整数). r_t 与 r_t^o 的关系用数学式子表示如下 [176]

$$r_t^o=\begin{cases}0 & \text{以概率 } \pi \\ r_t & \text{以概率}(1-\pi)^2 \\ r_t+r_{t-1} & \text{以概率}(1-\pi)^2\pi \\ r_t+r_{t-1}+r_{t-2} & \text{以概率}(1-\pi)^2\pi^2 \\ \vdots & \vdots \\ \sum_{i=0}^{k} r_{t-i} & \text{以概率}(1-\pi)^2\pi^k \\ \vdots & \vdots \end{cases} \tag{5.1}$$

这些概率很容易理解. 例如，$r_t^o=r_t$ 当且仅当 t 时刻与 $t-1$ 时刻都有交易发生；$r_t^o=r_t+r_{t-1}$，当且仅当 t 时刻与 $t-2$ 时刻都有交易发生，但是 $t-1$ 时刻无交易发生；$r_t^o=r_t+r_{t-1}+r_{t-2}$，当且仅当 t 时刻与 $t-3$ 时刻都有交易发生，但是 $t-1$ 时刻与 $t-2$ 时刻无交易发生；等等. 如所料想，全部的概率为 1，由下式给出

$$\pi+(1-\pi)^2[1+\pi+\pi^2+\cdots]=\pi+(1-\pi)^2\frac{1}{1-\pi}=\pi+1-\pi=1$$

我们准备考虑观测的收益率序列 $\{r_t^o\}$ 的矩方程. 首先，r_t^o 的期望是

$$
\begin{aligned}
E(r_t^o) &= (1-\pi)^2 E(r_t) + (1-\pi)^2 \pi E(r_t + r_{t-1}) + \cdots \\
&= (1-\pi)^2 \mu + (1-\pi)^2 \pi 2\mu + (1-\pi)^2 \pi^2 3\mu + \cdots \\
&= (1-\pi)^2 \mu [1 + 2\pi + 3\pi^2 + 4\pi^3 + \cdots] \\
&= (1-\pi)^2 \mu \frac{1}{(1-\pi)^2} = \mu
\end{aligned} \tag{5.2}
$$

在上面的推导中，我们利用了结果：$1+2\pi+3\pi^2+4\pi^3+\cdots=\frac{1}{(1-\pi)^2}$. 下一步，对于 r_t^o 的方差，我们利用 $\mathrm{Var}(r_t^o)=E[(r_t^o)^2]-[E(r_t^o)]^2$，以及

$$
\begin{aligned}
E(r_t^o)^2 &= (1-\pi)^2 E[(r_t)^2] + (1-\pi)^2 \pi E[(r_t + r_{t-1})^2] + \cdots \\
&= (1-\pi)^2 [(\sigma^2 + \mu^2) + \pi(2\sigma^2 + 4\mu^2) + \pi^2(3\sigma^2 + 9\mu^2) + \cdots] \qquad (5.3) \\
&= (1-\pi)^2 \{\sigma^2 [1 + 2\pi + 3\pi^2 + \cdots] + \mu^2 [1 + 4\pi + 9\pi^2 + \cdots]\} \qquad (5.4) \\
&= \sigma^2 + \mu^2 [\frac{2}{1-\pi} - 1] \qquad (5.5)
\end{aligned}
$$

在方程(5.3)中，我们利用了在 r_t 是前后独立的假定下，

177

$$
E\left(\sum_{i=0}^{k} r_{t-i}\right)^2 = \mathrm{Var}\left(\sum_{i=0}^{k} r_{t-i}\right) + \left[E\left(\sum_{i=0}^{k} r_{t-i}\right)\right]^2 = (k+1)\sigma^2 + [(k+1)\mu]^2.
$$

利用与方程(5.2)类似的方法，我们能够证明方程(5.4)的第一项简化为 σ^2. 对于方程(5.4)的第二项，我们利用恒等式

$$
1 + 4\pi + 9\pi^2 + 16\pi^3 + \cdots = \frac{2}{(1-\pi)^3} - \frac{1}{(1-\pi)^2},
$$

此恒等式可以如下推出：令

$$
H = 1 + 4\pi + 9\pi^2 + 16\pi^3 + \cdots \quad \text{且} \; G = 1 + 3\pi + 5\pi^2 + 7\pi^3 + \cdots.
$$

那么$(1-\pi)H=G$ 且

$$
\begin{aligned}
(1-\pi)G &= 1 + 2\pi + 2\pi^2 + 2\pi^3 + \cdots \\
&= 2(1 + \pi + \pi^2 + \cdots) - 1 = \frac{2}{1-\pi} - 1.
\end{aligned}
$$

因此，从方程(5.2)与方程(5.5)中可以得出

$$
\mathrm{Var}(r_t^o) = \sigma^2 + \mu^2\left[\frac{2}{1-\pi} - 1\right] - \mu^2 = \sigma^2 + \frac{2\pi\mu^2}{1-\pi}. \tag{5.6}
$$

下一步考虑$\{r_t^o\}$的 1 步延迟自协方差. 这里我们利用

$$
\mathrm{Cov}(r_t^o, r_{t-1}^o) = E(r_t^o r_{t-1}^o) - E(r_t^o)E(r_{t-1}^o) = E(r_t^o r_{t-1}^o) - \mu^2.
$$

则这个问题简化为寻找 $E(r_t^o r_{t-1}^o)$. 注意到如果 t 时刻无交易，或者 $t-1$ 时刻无交易，或者 t 时刻与 $t-1$ 时刻都无交易，则 $r_t^o r_{t-1}^o=0$. 因此，我们得到

$$r^o r_{t-1}^o = \begin{cases} 0 & \text{以概率 } 2\pi - \pi^2 \\ r_t r_{t-1} & \text{以概率}(1-\pi)^3 \\ r_t(r_{t-1}+r_{t-2}) & \text{以概率}(1-\pi)^3\pi \\ r_t(r_{t-1}+r_{t-2}+r_{t-3}) & \text{以概率}(1-\pi)^3\pi^2 \\ \vdots & \vdots \\ r_t\left(\sum_{i=1}^{k} r_{t-i}\right) & \text{以概率}(1-\pi)^3\pi^{k-1} \\ \vdots & \vdots \end{cases} \tag{5.7}$$

全部的概率和也是 1. 为了理解前面的结果，注意到 $r_t^o r_{t-1}^o = r_t r_{t-1}$ 当且仅当在时刻 $t-2$，$t-1$ 和 t 有三个连续的交易. 利用方程(5.7)以及对 $j>0$，$E(r_t r_{t-j}) = E(r_t)E(r_{t-j}) = \mu^2$，我们得到

$$E(r_t^o r_{t-1}^o) = (1-\pi)^3 \left\{ E(r_t r_{t-1}) + \pi E[r_t(r_{t-1}+r_{t-2})] + \pi^2 E\left[r_t\left(\sum_{i=1}^{3} r_{t-i}\right)\right] + \cdots \right\}$$

$$= (1-\pi)^3 \mu^2 [1 + 2\pi + 3\pi^2 + \cdots] = (1-\pi)\mu^2.$$

178

$\{r_t^o\}$的 1 步延迟自协方差为

$$\mathrm{Cov}(r_t^o, r_{t-1}^o) = -\pi\mu^2. \tag{5.8}$$

假定 μ 不为零，则非同步交易引起的 r_t^o 的负 1 步延迟自相关系数由下式给出

$$\rho_1(r_t^o) = \frac{-(1-\pi)\pi\mu^2}{(1-\pi)\sigma^2 + 2\pi\mu^2}.$$

一般地，我们可以推广前面的结果，并且证明

$$\mathrm{Cov}(r_t^o, r_{t-j}^o) = -\mu^2\pi^j, j \geqslant 1.$$

1 步延迟 ACF 的大小依赖于 μ、π 与 σ 的选择，并且可以是实实在在的不可忽视的. 这样，当 $\mu \neq 0$时，非同步交易就导致了观测到的证券收益率序列之间的负自相关性.

前面的讨论可以推广到一个包含 N 种证券的资产组合的收益率序列；见 Campbell，Lo 和 MacKinlay(1997，第 3 章). 在时间序列文献中，非同步交易对于单个证券收益率的效应等价于关于一个时间序列的随机的时间积累，此时交易的概率 π 控制了积累的机制.

5.2 买卖报价差

在某些股票交易所(如 NYSE)，做市商(market maker) 在促进交易方面起了非常重要的作用. 他们提供了市场流动性(market liquidity)：每当公众有买卖的愿望时，他们随时都准备好进行买或卖. 市场流动性是指能快速地、匿名地、几乎没有价格影响地买卖相当数量证券的能力. 作为提供流动性的回报，交易所赋予做市商对证券的买卖双方传递不同价格的专利权. 他们以标价 P_b 购买，以更高的叫价 P_a 卖出（对公众来说，P_b 是卖出价格，P_a 是买入价格）. 价格差$P_a - P_b$称为买卖报价差(bid-ask spread)，这是做市商获得报酬的主要来源. 买卖报价差一般比较小，也就是一两个最小升降档.

买卖报价差尽管数量上比较小，但是它的存在对于资产收益率时间序列的性质有几个重要的影响. 我们这里主要讨论买卖报价弹性，即买卖报价差引起的资产收益率的1步延迟负序列相关. 考虑Roll(1984)的简单模型，假定观测到的资产市场价格P_t满足

179

$$P_t = P_t^* + I_t \frac{S}{2} \tag{5.9}$$

其中$S=P_a-P_b$表示买卖报价差，P_t^*表示一个无摩擦市场中资产在t时刻的基本价值，$\{I_t\}$是一个独立的、服从等概率二项分布的随机变量序列(即以概率0.5，$I_t=1$；以概率0.5，$I_t=-1$). I_t可以解释为一个指令型的示性变量，取1表示买方发动的交易；取-1表示卖方发动的交易. 换一种说法，模型可以写为

$$P_t = P_t^* + \begin{cases} +S/2 & \text{以概率 } 0.5 \\ -S/2 & \text{以概率 } 0.5 \end{cases}$$

如果P_t^*不变，那么价格变化的观测过程为

$$\Delta P_t = (I_t - I_{t-1}) \frac{S}{2}. \tag{5.10}$$

在方程(5.9)对于I_t的假定下，我们有$E(I_t)=0$，$\mathrm{Var}(I_t)=1$，从而可以得到$E(\Delta P_t)=0$，而且

$$\mathrm{Var}(\Delta P_t) = S^2/2, \tag{5.11}$$

$$\mathrm{Cov}(\Delta P_t, \Delta P_{t-1}) = -S^2/4, \tag{5.12}$$

$$\mathrm{Cov}(\Delta P_t, \Delta P_{t-j}) = 0, \quad j > 1. \tag{5.13}$$

因此，ΔP_t的自相关函数为

$$\rho_j(\Delta P_t) = \begin{cases} -0.5 & \text{若 } j = 1, \\ 0 & \text{若 } j > 1. \end{cases} \tag{5.14}$$

这样买卖报价差就导致观测到的价格变化序列的1步延迟负相关，这在金融文献中一般称为买卖报价弹性. 对此，我们可以从下面的直观意义上来理解. 假定基本价格P_t^*等于$(P_a+P_b)/2$，那么P_t就有P_a与P_b两种取值. 如果前面观测到的价格为P_a(较高的值)，则当前观测到的价格要么不变，要么为P_b，从而ΔP_t要么为0，要么为$-S$；然而，如果前面观测到的价格为P_b(较低的值)，则ΔP_t要么为0，要么为S. ΔP_t的1步延迟负相关变得很明显. 但是，买卖报价差并不会引起任何超过1步延迟的序列相关.

一个更加现实的表示是假定P_t^*服从随机游走，满足$\Delta P_t^* = P_t^* - P_{t-1}^* = \epsilon_t$，这形成了一个均值为0，方差为$\sigma^2$的独立同分布的随机变量序列，并且$\{\epsilon_t\}$与$\{I_t\}$独立. 在这种情形，$\mathrm{Var}(\Delta P_t)=\sigma^2+S^2/2$，但是$\mathrm{Cov}(\Delta P_t, \Delta P_{t-j})$是不变的，从而

180

$$\rho_1(\Delta P_t) = \frac{-S^2/4}{S^2/2+\sigma^2} \leqslant 0.$$

虽然ΔP_t的1步延迟自相关的大小减小了，但是当$S=P_a-P_b>0$时，负效应还是存在的. 金融中，研究买卖报价差的组成是比较有趣的，感兴趣的读者可以参考Campbell，Lo和

MacKinlay(1997) 及其参考文献.

买卖报价差的效应在组合收益率与多元金融时间序列中也是存在的. 考虑二元的情形，用 $I_t=(I_{1t}, I_{2t})'$表示二元指令型示性变量，其中 I_{1t}是第一种证券的示性变量，I_{2t}是第二种证券的示性变量. 如果 I_{1t}与 I_{2t}是同步相关的，则买卖报价差可以引起负的 1 步延迟交叉相关.

5.3 交易数据的经验特征

令 t_i 表示资产的第 i 次交易发生的时刻，它是从午夜开始以秒为单位测量的日历时间. 与交易相关的有几个变量如：交易价格、交易量、盛行的买卖报价等等. t_i 的集合与相关的度量一般称为交易数据，这些数据有几个重要的特征，这些特征是当观测值随时间加总时不存在的，我们在下面给出一些特征.

1)不等间隔的时间区间：交易(如交易所里面的股票交易)一般不是在等间隔的时间区间上发生的，这样观测到的资产的交易价格并不形成一个等间隔的时间序列. 从而交易间的时间持续期变得非常重要，并且可能包含了关于市场微观结构(例如交易强度)的有用信息.

2)离散取值的价格：资产从一个交易到另外一个交易的价格变化只在最小变动价位的倍数中发生. 在 NYSE 中，最小变动价位 在 1997 年 6 月 24 日以前是 1/8 美元，2001 年 1 月 29 日以前是 1/16 美元. 在 2001 年 1 月 29 日，所有的 NYSE 和 AMEX 股票都开始以小数为单位进行交易，因此，价格在交易数据中是离散取值的变量. 在某些市场中，价格的变化也可能受监管者设定的条件的限制.

3)日周期或者日模式的存在：在正常交易条件下，交易活动能够展示周期模式. 举例说，在 NYSE 中，开盘与收盘时刻的交易比较频繁，而中午时间交易比较少，导致了“U”型的交易强度. 因此，交易之间的时间持续期亦呈现日循环模式.

4)一秒钟的多重交易：多重交易，甚至是具有不同价格的多重交易同时发生是可能的，部分由于时间是以秒来测量的，而在大量交易期间，这种时间长度可能还是太长. [181]

为了描述这些特征，首先考虑从 1990 年 11 月 1 日到 1991 年 1 月 31 日的 IBM 交易数据，这些数据来自于 TORQ(即交易、指令、报告和报价)数据集；见 Hasbrouck(1992). 共有 63 个交易日，60 328 次交易. 为了简化讨论，我们忽略了交易日间的价格变化，集中讨论发生在美国东部时间早上 9：30 至下午4：00的正常交易时间里的交易. 众所周知，头天晚上的股票收益率与当天的收益率有显著不同；见 Stoll 和 Whaley(1990) 及其参考文献. 表 5-1 以百分比形式给出了价格变化的频率，其中最小变动价位为 1/8 美元＝0. 125 美元. 从表中我们做出以下结论：

1)大约 2/3 的日交易是没有价格变化的；

2)一个最小变动价位的价格变化接近于日交易的 29%；

3)只有 2. 6%的交易与两个最小变动价位的价格变化相关；

4)只有 1. 3%的交易导致三个或者更多个最小变动价位的价格变化；

5)正负价格变化的分布是渐近对称的.

表 5-1 IBM 股票从 1990 年 11 月 1 日到 1991 年 1 月 31 日以最小变动价位的倍数变化的价格变化频率

数量(tick)	≤−3	−2	−1	0	1	2	≥3
百分比	0.66	1.33	14.53	67.06	14.53	1.27	0.63

下面考虑每五分钟时间间隔上的交易量，用 x_t 来表示此序列. 也就是，x_1 表示在 1990 年 11 月 1 日美国东部时间，IBM 从上午 9∶30 至 9∶35 的交易量；x_2 表示从 9∶35 至9∶40的交易量，等等. 忽略交易日之间的时间间隙. 图 5-1a 画出了 x_t 的时间图，图 5-1b 描述了 x_t 延迟 1 阶至 260 阶的样本 ACF. 尤其有趣的是 ACF 的周期为 78 的循环模式，这正好是一个交易日内每 5 分钟间隔的数量. 这样，交易量呈现出日模式. 为了更进一步描述日交易模式，图 5-2 画出了 63 天的每 5 分钟的平均交易量，共有 78 个这样的平均数. 这个图形呈现出“微笑”形状或者“U”的形状，表明在市场开盘与收盘的时候交易多，而午时交易少的特点.

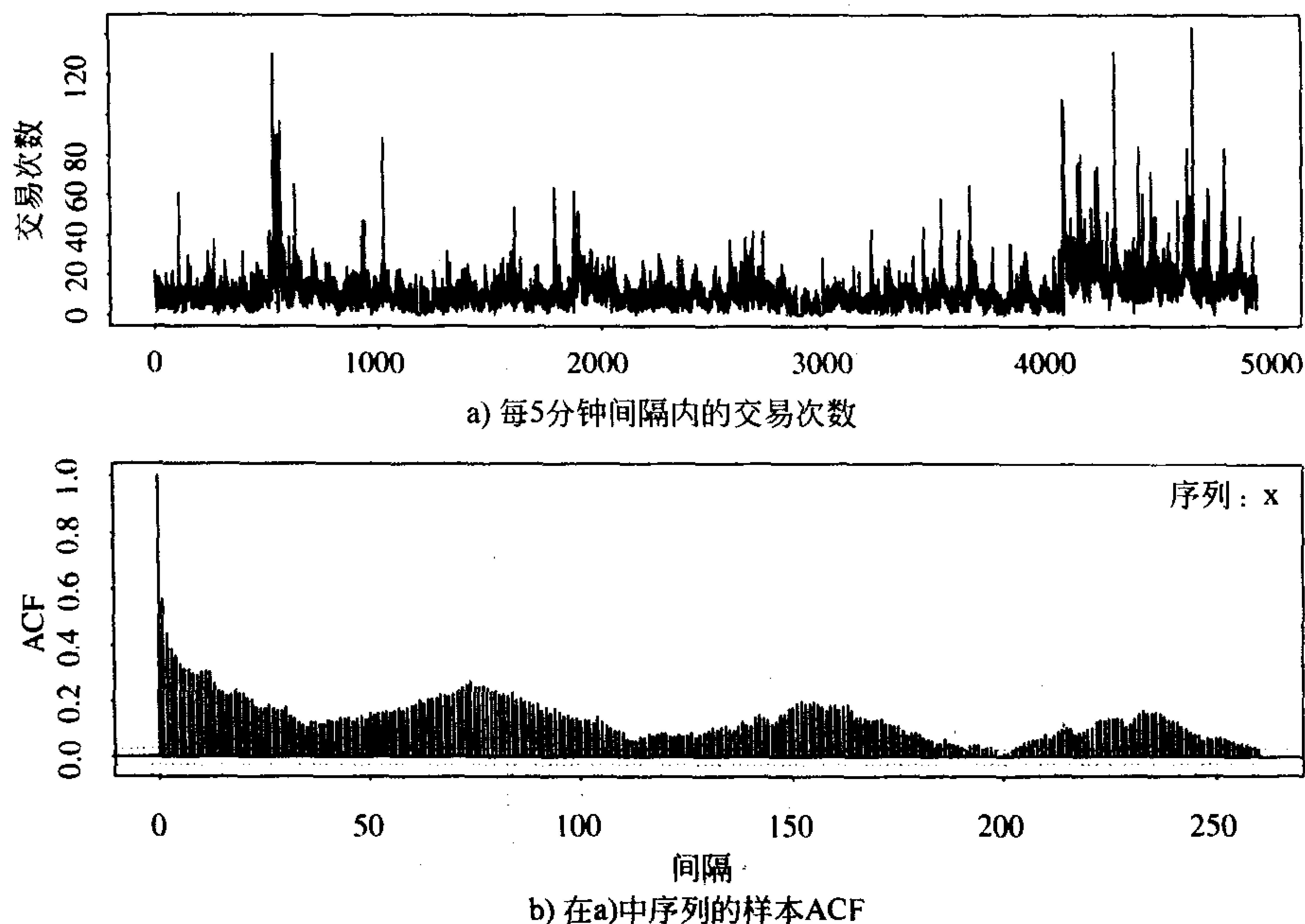


图 5-1 IBM 从 1990 年 11 月 1 日至 1991 年 1 月 31 日的日交易数据

我们着重讨论在一个交易日中的正常交易时间发生的交易，在数据中总共有 59 838 个时间间隔. 这些间隔称为交易之间的日持续期. 对 IBM 股票，有 6 531 个零时间间隔. 也就是，从 1990 年 11 月 1 日至 1991 年 1 月 31 日之间的 63 个交易日的正常交易时间，每秒钟的多重交易发生了 6 531 次，大约是 10.91%. 在这些多重交易中，有 1 002 个有着不同的价格，这大约是全部日交易量的 1.67%. 因此，多重交易(即零息持续期)可能会成为交易之间的时间持续期统计模拟中的一个问题.

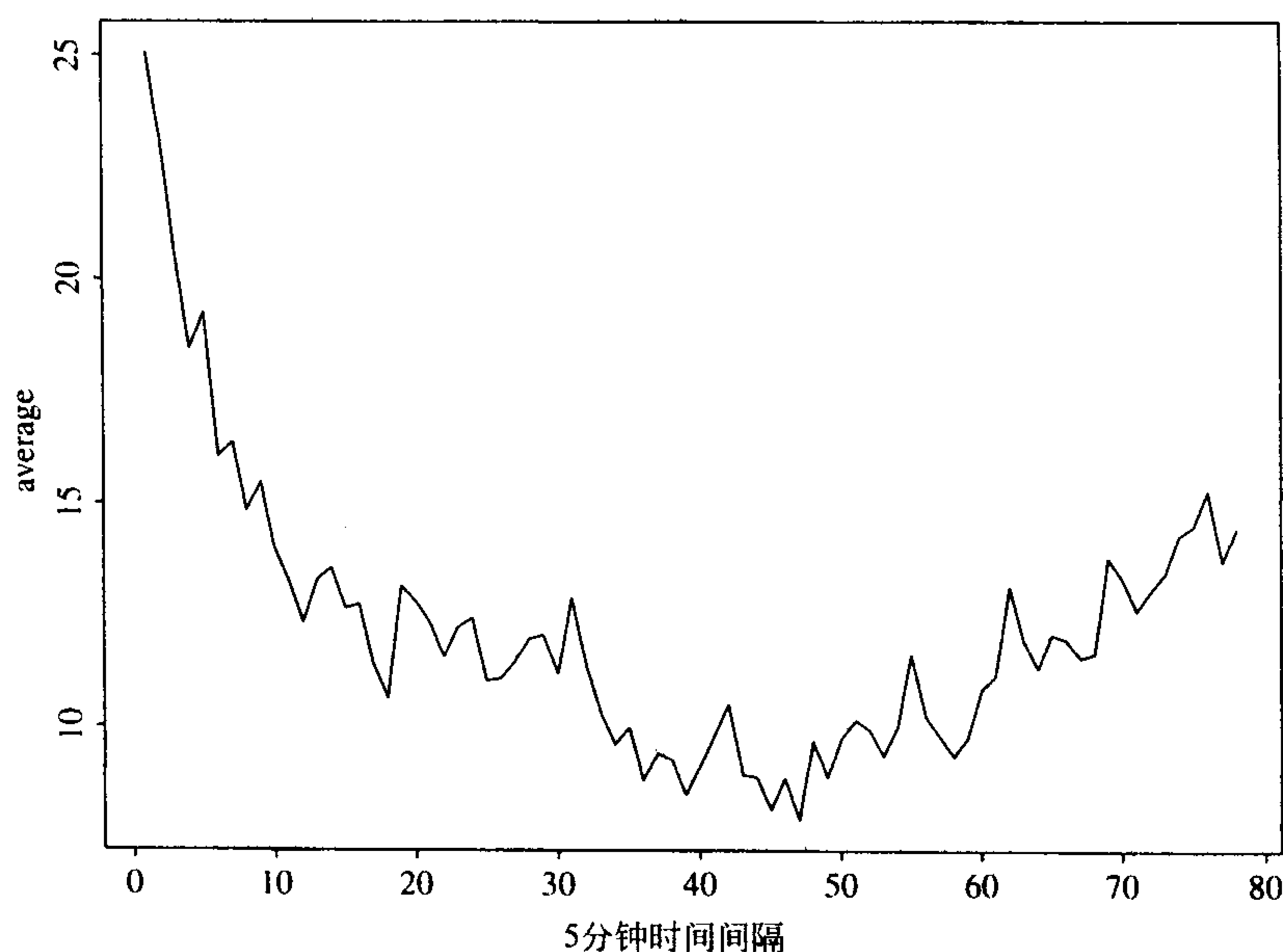


图 5-2 每 5 分钟间隔的平均交易次数的时间图. IBM 股票从 1990 年 11 月 1 日至 1991 年 1 月 31 日共 63 个交易日的平均，共 78 个观测值

表 5-2 提供了价格运动的两种分类方式. 这里，价格运动被分成"上升"、"不变"、"下降"三类，分别表示为"+"、"0"和"-". 此表给出了样本在两个连续交易之间(从第 $i-1$ 个交易至第 i 交易)的价格运动. 由表中交易对交易的数据显示：

1)连续的价格递增或者递减的情形是相当少的，分别是 441/59 837=0.74%，410/59 837=0.69%；

2)从表的第一行看出，有轻微的从"上升"到"不变"而不是到"下降"的移动；

3)价格保持"不变"的倾向很高；

4)从第 3 行可以看出，从"下降"到"上升"或者"不变"的概率大约是相同的.

表 5-2

第($i-1$)次交易	第 i 次交易			边际
	"+"	"0"	"-"	
"+"	441	5 498	3 948	9 887
"0"	4 867	29 779	5 473	40 119
"-"	4 580	4 841	410	9 831
边际	9 888	40 118	9 831	59 837

注：IBM 股票相邻两日内交易的价格运动的分类. 价格运动分成"上升"、"不变"、"下降"三类. 数据的时间区间为 11/01/90～1/31/91.

前面提到的第一个观测就是买卖报价弹性的一个清楚的说明，显示了日交易数据的价格逆转. 为了确认这个现象，我们考虑价格运动的一个方向序列 D_i，在第 i 次交易中，对于价格运

动“上升”、“不变”、“下降”三种情况，D_i 分别取+1、0、−1 三个值. $\{D_i\}$的 ACF 在延迟为 1 时的取值−0.389 是一个单峰，这对于样本大小为 59 837 的序列来说是高度显著的，证实了连续交易中的价格逆转.

作为第二个解释，我们考虑由 TAQ 数据库得到的 IBM 股票在 1999 年 12 月的交易数据. 正常的交易时间是从美国东部时间早上 9：30 至下午 4：00，12 月 31 日这天除外，因为这天市场在下午 13：00 关闭. 与 1990～1991 年的数据比较，发生了两个重要的变化. 首先，日交易次数增加了六倍，仅在 1999 年 12 月就发生了 134 120 次交易. 交易强度的增加也给每秒钟的多重交易增加了机会. 零时间持续期的交易百分比增长了两倍，到了 22.98%. 更极端的是，1999 年 12 月 3 日在给定的一秒钟内发生 42 次交易的情形出现了两次. 第二，价格运动的最小变动价位是 1/16 美元＝0.062 5 美元，而不是 1/8 美元. 最小变动价位的变化降低了买卖报价差. 图 5-3 给出了新样本的日交易次数. 图 5-4a 是以秒为单位测量的交易之间的时间持续期的时间图，图 5-4b 是用最小变动价位 1/16 美元的倍数测量的连续日交易的价格变化图. 如所料想，图 5-3 与图 5-4a 很清楚地说明了日交易次数与交易之间的时间间隔之间的逆转关系. 图 5-4b 描述了 1999 年 12 月 3 日 IBM 股票的两个不寻常的价格运动：其一是先下降了 63 个价位，紧接着上涨了 64 个价位；其二是下降 68 个价位后，紧接着上涨了 68 个价位. 在一个交易日的交易中，像这样不寻常的价格变化是很少发生的.

对于在正常的交易时刻中记录的交易，在 133 475 个交易中有 61 149 个交易没有价格变化，这大约是 45.8%，比 1990 年 11 月到 1991 年 1 月这个时间段上的没有价格变化的交易数少很多. 看上去是减少最小变动价位增加了价格变化的机会. 表 5-3 给出了与价格变化相关的交易数的百分比. 价格运动仍然近似地关于零对称. 日交易中很大的价格变化仍然相对很少.

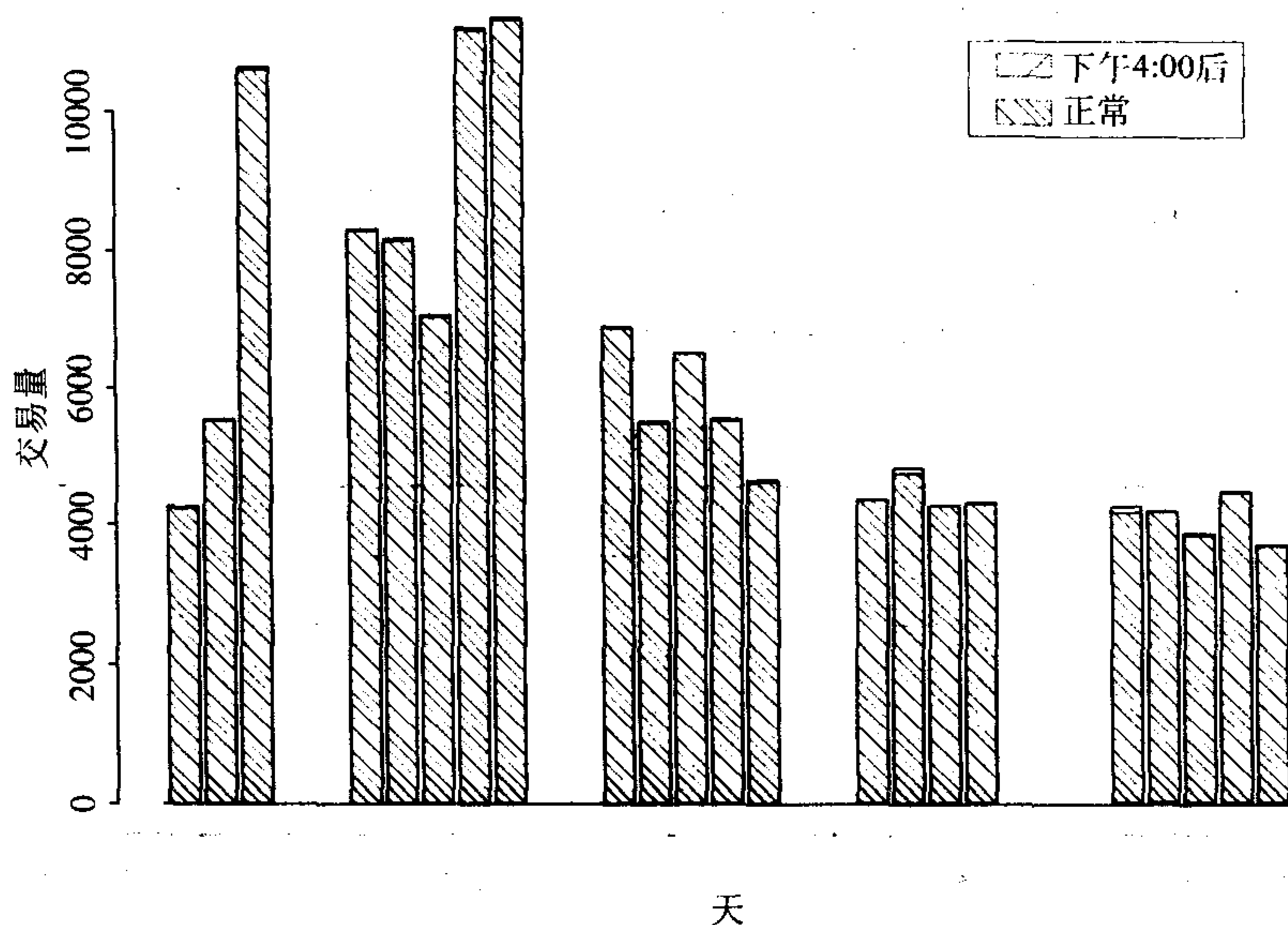
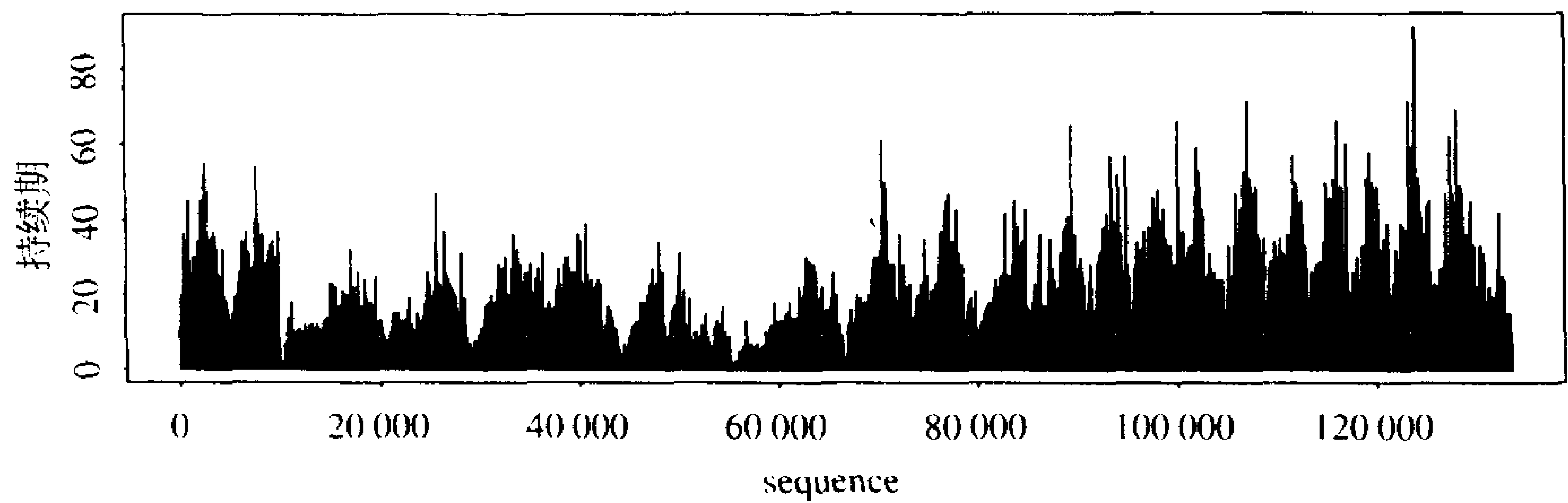
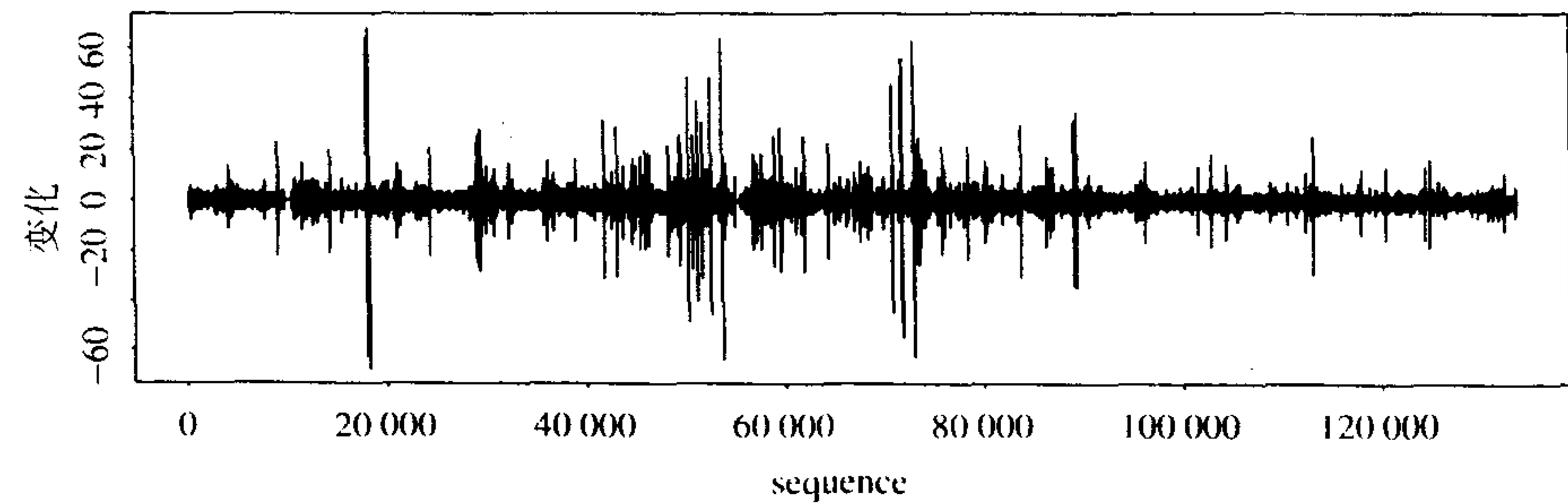


图 5-3 1999 年 12 月 IBM 的交易数据. 每个交易日的交易数，其中有的长条加上了标记时间为当天下午4：00以后发生(after-hours)的交易数



a) 交易之间时间持续期的时间图



b) 以最小变动价位1/16美元的倍数测量的接连两个交易的价格变化量的时间图，只包含正常交易时间的数据

图 5-4 1999 年 12 月 IBM 的交易数据

表 5-3

大小	1	2	3	4	5	6	7	>7
				(a)上升运动				
百分比	18.03	5.80	1.79	0.66	0.25	0.15	0.09	0.32
				(b)下降运动				
百分比	18.24	5.57	1.79	0.71	0.24	0.17	0.10	0.31

注：与 1999 年 12 月交易的 IBM 股票的价格变化相关的日交易百分比. 在正常的交易时刻，没有价格变化的交易的百分比为 45.8%，全部的交易量为 133 475. 这个大小是以最小变动价位 1/16 美元的倍数测量的.

注释： 高频数据的保存记录并不像以较低频率抽取的观测值那么好，因此，在高频数据分析中，很有必要进行数据清理. 对于交易数据，很多方式都可以造成观测值的缺失，从而对某些交易而言，精确交易时间的正确程度是值得怀疑的. 例如，记录的交易时间可能超过了美国东部时间的下午 4：00，甚至在下午4：00后的交易开始之前. 怎样处理这些观测值值得仔细研究. 数据清理的一个适当方法要求对市场操作的方式有着很深的理解. 因而，清楚准确地鉴定数据清理中使用的方法是非常重要的，在作出推断之前，必须考虑这些方法.

185~186

再令 t_i 表示从午夜开始以秒为单位测量的第 i 次交易发生的日历时间，P_{t_i} 表示交易价格. 从第 $i-1$ 次交易至第 i 次交易的价格变化为 $y_i \equiv \Delta P_{t_i} = P_{t_i} - P_{t_{i-1}}$，时间持续期为 $\Delta t_i = t_i - t_{i-1}$. 这里理解为，$\Delta t_i$ 与 y_i 的下标 i 表示交易的时间序列，而不是日历时间. 在下面的讨论中，我们主要考虑 y_i 与 Δt_i 的单个模型与联合模型.

5.4 价格变化模型

离散化与集中于"无变化"这两个性质使得我们对日价格变化建模比较困难. Campbell，Lo 和 Mackinlay(1997) 讨论了文献中已经提出的几种经济计量模型. 在这里，我们提及两种利用解释变量来研究日交易价格变化的模型. 第一个模型是 Hauseman，Lo 和 Mackinlay (1992) 在研究交易数据的价格运动中使用的顺序概率值模型. 第二个模型最近由 McCulloch 和 Tsay(2000) 提出，此模型是 Rydberg 和 Shephard(1998) 提出的模型的一个简便形式，也可以参考 Ghysels(2000).

5.4.1 顺序概率值模型

令 y_i^* 表示所研究的资产的不能观测到的价格变化(也就是，$y_i^* = P_{t_i}^* - P_{t_{i-1}}^*$，其中 P_t^* 表示资产在 t 时刻的真实价格. 顺序概率值模型假定 y_i^* 是一个连续的随机变量，服从以下模型

$$y_i^* = \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i \tag{5.15}$$

其中 $\boldsymbol{x}_i$ 表示 t_{i-1}时刻可以得到的解释变量的 p 维行向量，$\boldsymbol{\beta}$是一个 $p\times 1$ 参数向量，$E(\varepsilon_i \mid \boldsymbol{x}_i)=0$，$\mathrm{Var}(\varepsilon_i \mid \boldsymbol{x}_i)=\sigma_i^2$，并且对于 $i \neq j$，$\mathrm{Cov}(\varepsilon_i, \varepsilon_j)=0$. 假定条件方差 σ_i^2 是解释变量 $\boldsymbol{w}_i$ 的正函数，即

$$\sigma_i^2 = g(\boldsymbol{w}_i) \tag{5.16}$$

其中 $g(.)$是一个正函数. 对于金融交易数据而言，$\boldsymbol{w}_i$ 可能包含了时间间隔 $t_i - t_{i-1}$ 以及某些条件异方差变量. 特别地，也可以假定 ε_i 在给定 $\boldsymbol{x}_i$ 和 $\boldsymbol{w}_i$ 下的条件分布是高斯的.

假定观测到的价格变化 y_i 有 k 个可能的取值，理论上 k 可以是无穷大，但必须是可数的. [187] 实际应用中，k 是有限的，可能涉及到将几种类别组合成单个值. 例如，在表 5-1 中我们有 $k=7$,其中第一个值"-3 价位"意味着价格的变化为-3 个价位或者更低，我们将 k 个可能的取值表示为$\{s_1, s_2, \cdots, s_k\}$，则顺序概率值模型假定 y_i 与 y_i^* 之间的关系为

$$y_i = s_j, \quad 若\ \alpha_{j-1} < y_i^* \leqslant \alpha_j, \quad j = 1,\cdots,k \tag{5.17}$$

其中 α_j 是实数，满足$-\infty=\alpha_0<\alpha_1<\cdots<\alpha_{k-1}<\alpha_k=\infty$. 在条件高斯分布的假定下，我们有

$$\begin{aligned} P(y_i = s_j \mid \boldsymbol{x}_i, \boldsymbol{w}_i) &= P(\alpha_{j-1} < \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i \leqslant \alpha_j \mid \boldsymbol{x}_i, \boldsymbol{w}_i) \\ &= \begin{cases} P(\boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i \leqslant \alpha_1 \mid \boldsymbol{x}_i, \boldsymbol{w}_i) & 如果\ j=1, \\ P(\alpha_{j-1} < \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i \leqslant \alpha_j \mid \boldsymbol{x}_i, \boldsymbol{w}_i) & 如果\ j=2,\cdots,k-1, \\ P(\alpha_{k-1} < \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i \mid \boldsymbol{x}_i, \boldsymbol{w}_i) & 如果\ j=k, \end{cases} \end{aligned}$$

$$= \begin{cases} \Phi\left[\dfrac{\alpha_1 - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right] & \text{若 } j=1, \\ \Phi\left[\dfrac{\alpha_j - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right] - \Phi\left[\dfrac{\alpha_{j-1} - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right] & \text{若 } j=2,\cdots,k-1, \\ 1 - \Phi\left[\dfrac{\alpha_{k-1} - \boldsymbol{x}_i\boldsymbol{\beta}}{\sigma_i(\boldsymbol{w}_i)}\right], & \text{若 } j=k, \end{cases} \tag{5.18}$$

其中 $\Phi(x)$是标准正态随机变量的累积分布函数在 x 点的取值，并且我们用$\sigma_i(\boldsymbol{w}_i)$表示 σ_i^2 是 $\boldsymbol{w}_i$ 的正函数. 从定义中可以看出，顺序概率值模型是由未观测到的连续随机变量所驱动的. 观测值都有一个自然序号，可以认为是代表所在过程的类别.

顺序概率值模型包含了参数$\boldsymbol{\beta}$，$\alpha_i(i=1, \cdots, k-1)$以及方程(5.16)中条件方差函数 $\sigma_i(\boldsymbol{w}_i)$ 包含的参数. 这些参数可以通过最大似然估计或者是马尔可夫链蒙特卡罗(MCMC)方法来估计.

例 5.1　Hauseman，Lo 和 Mackinlay(1992) 对 1988 年 100 多种股票的交易数据应用顺序概率值模型，这里，我们仅仅描述他们对于 IBM 的结果，共有206 794个交易. 价格变化 y_i 的样本均值(标准差)为$-0.001\,0(0.753)$，时间持续期 Δt_i 的样本均值(标准差)为27.21(34.13)，买卖报价差的样本均值(标准差) 为 1.947 0(1.462 5). 买卖报价差是用最小变动价位来测量的. 运用的模型对于价格运动有 9 个分类，而且指定函数为

$$\boldsymbol{x}_i\boldsymbol{\beta} = \beta_1 \Delta t_i^* + \sum_{v=1}^{3} \beta_{v+1} y_{i-v} + \sum_{v=1}^{3} \beta_{v+4} \mathrm{SP5}_{i-v} + \sum_{v=1}^{3} \beta_{v+7} \mathrm{IBS}_{i-v} + \sum_{v=1}^{3} \beta_{v+10} [T_\lambda(V_{i-v}) \times \mathrm{IBS}_{i-v}] \tag{5.19}$$

188

$$\sigma_i^2(\boldsymbol{w}_i) = 1.0 + \gamma_1^2 \Delta t_i^* + \gamma_2^2 \mathrm{AB}_{i-1} \tag{5.20}$$

其中 $T_\lambda(V)=(V^\lambda-1)/\lambda$ 是指对于$\lambda\in[0, 1]$的 V 的 Box-Cox(1964)变换，解释变量定义如下：

- $\Delta t_i^* = (t_i - t_{i-1})/100$ 是第 $i-1$ 次交易与第 i 次交易之间的尺度修正的时间持续期，这里时间是以秒为单位测量的.
- AB_{i-1}是时刻 t_{i-1}流行的以最小变动价位计算的买卖报价差.
- $y_{i-v}(v=1, 2, 3)$是 t_{i-v}时刻以最小变动价位计算的价格变化的延迟值. 对于 $k=9$，以最小变动价位计算的价格变化的可能取值为$\{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$.
- $V_{i-v}(v=1, 2, 3)$是第 $i-v$ 次交易中的交易额(美元数)的延迟值，定义为第 $i-v$ 次交易的美元价格乘以交易股份数. 这里，美元数以百美元为单位.
- $\mathrm{SP5}_{i-v}(v=1, 2, 3)$是标准普尔 500 指数在与第 $i-v$ 次交易发生的月份最靠近的那个月到期的期货价格的每五分钟连续复合收益率，这里的收益率是用 t_{i-v}以前的最近取整时间的前一分钟和前五分钟记录的期货价格来计算的.
- $\mathrm{IBS}_{i-v}(v=1, 2, 3)$是一个示性变量，定义为

$$\mathrm{IBS}_{i-v} = \begin{cases} 1 & \text{当 } P_{i-v} > (P_{i-v}^a + P_{i-v}^b)/2 \\ 0 & \text{当 } P_{i-v} = (P_{i-v}^a + P_{i-v}^b)/2 \\ -1 & \text{当 } P_{i-v} < (P_{i-v}^a + P_{i-v}^b)/2 \end{cases}$$

其中 P_j^a，P_j^b 分别为 t_j 时刻的卖价与买价.

表 5-4 给出了参数估计以及它们的 t 比. 除了一个 t 比较小以外，其余 t 比都很大，显示了估计是高度显著的. 因为样本的容量很大，所以如此高的 t 比并不令人感到惊奇. 对于交易量很大的 IBM 股票而言，估计结果显示出如下结论：

1)边界的划分并不是等间隔的，但是关于 0 几乎是对称的.

2)交易的持续期 Δt_i 不仅影响方程(5.19)表示的 y_i 的条件均值，而且影响方程(5.20)表示的 y_i 的条件方差.

3)延迟价格变化的系数为负，并且是高度显著的，显示了价格逆转性质.

189 4)如所料想的，t_{i-1}时刻的买卖报价差显著地影响条件方差.

表 5-4 对 IBM 股票 1988 年的交易数据，方程(5.19)、(5.20)中的顺序概率值模型的参数估计，其中 t 表示 t 比

(a)概率值模型的边界分解

参数	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
估计	−4.67	−4.16	−3.11	−1.34	1.33	3.13	4.21	4.73
t	−145.7	−157.8	−171.6	−155.5	154.9	167.8	152.2	138.9

(b) 概率值模型的方程参数

参数	γ_1	γ_2	β_1 : Δt_i^*	β_2 : y_{-1}	β_3	β_4	β_5	β_6
估计	0.40	0.52	−0.12	−1.01	−0.53	−0.21	1.12	−0.26
t	15.6	71.1	−11.4	−135.6	−85.0	−47.2	54.2	−12.1
参数	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	
估计	0.01	−1.14	−0.37	−0.17	0.12	0.05	0.02	
t	0.26	−63.6	−21.6	−10.3	47.4	18.6	7.7	

5.4.2 分解模型

对价格变化建模的另外一个方法是将价格变化分解为三个组成部分，然后运用每一部分的条件确定，可以参见 Rydberg 和 Shephard(1998). 这三个部分分别为价格变化的示性量，有价格变化时价格变化的方向，以及价格变化发生时变化的大小. 具体地，在第 i 次交易中的价格变化可以写为

$$y_i \equiv P_{t_i} - P_{t_{i-1}} = A_i D_i S_i. \tag{5.21}$$

其中 A_i 是一个二值变量，定义为

$$A_i = \begin{cases} 1, & \text{如果第 } i \text{ 次交易中有价格变化,} \\ 0, & \text{如果第 } i \text{ 次交易中无价格变化.} \end{cases} \tag{5.22}$$

D_i 也是一个离散变量，它表示有价格变化的时候，价格变化的方向，即

$$D_i \mid (A_i = 1) = \begin{cases} 1, & \text{第 } i \text{ 次交易中价格上升,} \\ -1, & \text{第 } i \text{ 次交易中价格下降.} \end{cases} \tag{5.23}$$

其中 $D_i \mid (A_i=1)$ 意思是说 D_i 是在 $A_i=1$ 的条件下定义的. 当第 i 次交易中有价格变化时，S_i 是指以最小变动价位的价格变化的大小；当第 i 次交易中没有价格变化时，$S_i=0$. 当有价格变化时，S_i 是一个正的整型值随机变量.

注意到 $A_i=0$ 时，不需要 D_i，并且在分解中有一个自然的次序. 只有当 $A_i=1$ 时，D_i 才有定义；只有当 $A_i=1$ 并且给定 D_i 的条件下，S_i 才有意义. 分解中模型的确定利用了这个次序. [190]

令 F_i 表示第 i 次交易中可以得到的信息集合. F_i 的元素为 Δt_{i-j}，A_{i-j}，D_{i-j} 以及 $S_{i-j}(j\geqslant 0)$. 在模型(5.21)下，价格变化的演变可以分解为

$$P(y_i \mid F_{i-1}) = P(A_iD_iS_i \mid F_{i-1}) = P(S_i \mid D_i, A_i, F_{i-1})P(D_i \mid A_i, F_{i-1})P(A_i \mid F_{i-1}) \tag{5.24}$$

因为 A_i 是一个二进制变量，所以只要考虑随时间演变的概率 $p_i=P(A_i=1)$ 就足够了. 我们假定

$$\ln\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i\boldsymbol{\beta} \quad \text{或} \quad p_i = \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}} \tag{5.25}$$

其中 $\boldsymbol{x}_i$ 是一个包含了 F_{i-1} 中的所有元素的有限维向量，$\boldsymbol{\beta}$ 是一个参数向量. 在 $A_i=1$ 的条件下，D_i 也是一个二进制变量，对 $\delta_i=P(D_i=1 \mid A_i=1)$，我们利用下面的模型

$$\ln\left(\frac{\delta_i}{1-\delta_i}\right) = \boldsymbol{z}_i\boldsymbol{\gamma} \quad \text{或} \quad \delta_i = \frac{e^{\boldsymbol{z}_i\boldsymbol{\gamma}}}{1+e^{\boldsymbol{z}_i\boldsymbol{\gamma}}} \tag{5.26}$$

其中 $\boldsymbol{z}_i$ 是一个包含了 F_{i-1} 中的所有元素的有限维向量，$\boldsymbol{\gamma}$ 是一个参数向量. 为了允许正负价格变化之间的对称性，我们假定

$$S_i \mid (D_i, A_i=1) \sim 1 + \begin{cases} g(\lambda_{u,i}) & \text{如果 } D_i = 1, A_i = 1 \\ g(\lambda_{d,i}) & \text{如果 } D_i = -1, A_i = 1 \end{cases} \tag{5.27}$$

其中 $g(\lambda)$ 是参数为 λ 的几何分布，并且参数 $\lambda_{j,i}$ 随时间的变化为

$$\ln\left(\frac{\lambda_{j,i}}{1-\lambda_{j,i}}\right) = \boldsymbol{w}_i\boldsymbol{\theta}_j \quad \text{或} \quad \lambda_{j,i} = \frac{e^{\boldsymbol{w}_i\boldsymbol{\theta}_j}}{1+e^{\boldsymbol{w}_i\boldsymbol{\theta}_j}}, \qquad j = u, d, \tag{5.28}$$

其中 $\boldsymbol{w}_i$ 指 F_{i-1} 中有限维的解释变量，$\boldsymbol{\theta}_j$ 是一个参数向量.

在方程(5.27)中，随机变量 x 服从几何分布 $g(\lambda)$，其概率质量函数为

$$p(x=m) = \lambda(1-\lambda)^m, \qquad m = 0,1,2,\cdots$$

我们在几何分布中加了 1，以便当发生价格变化的时候，价格的变化至少为一个最小变动价位. 在方程(5.28)中，为了确保 $\lambda_{j,i}\in[0,1]$，我们作了 logistic 变换. [191]

前面的指定将第 i 次交易分为三类：

1)无价格变化：$A_i=0$ 而且相应的概率为 $(1-p_i)$；

2)价格上升：$A_i=1$，$D_i=1$，相应的概率为 $p_i\delta_i$. 价格上升的大小是由 $1+g(\lambda_{u,i})$ 控制的.

3)价格下降：$A_i=1$，$D_i=-1$，相应的概率为 $p_i(1-\delta_i)$. 价格下降的大小是由 $1+g(\lambda_{d,i})$ 控制的.

对于 $j=1,2,3$，令 $I_i(j)$ 表示前面三类的示性变量，也就是，当第 j 类发生时，$I_i(j)=$

1，当第 j 类不发生时，$I_i(j)=0$. 方程(5.24)的对数似然函数为

$$\ln[P(y_i \mid F_{i-1})] = I_i(1)\ln[(1-p_i)] + I_i(2)[\ln(p_i) + \ln(\delta_i) + \ln(\lambda_{u,i}) + (S_i-1)\ln(1-\lambda_{u,i})] + I_i(3)[\ln(p_i) + \ln(1-\delta_i) + \ln(\lambda_{d,i}) + (S_i-1)\ln(1-\lambda_{d,i})]$$

并且全部的对数似然函数为

$$\ln[P(y_1,\cdots,y_n \mid F_0)] = \sum_{i=1}^{n}\ln[P(y_i \mid F_{i-1})] \tag{5.29}$$

这是参数$\boldsymbol{\beta}$，$\boldsymbol{\gamma}$，$\boldsymbol{\theta}_u$ 以及$\boldsymbol{\theta}_d$ 的一个函数.

例 5.2 我们通过解释 IBM 股票从 1990 年 11 月 1 日至 1991 年 1 月 31 日的日交易来说明分解模型. 共有 63 个交易日，在正常的交易时间发生了 59 838 次日内交易. 运用的解释变量为

1) A_{i-1}：上一次交易的行动示性变量(即在一个交易日内的第 $i-1$ 次交易).

2) D_{i-1}：上一次交易的方向示性变量.

3) S_{i-1}：上一次交易的大小.

4) V_{i-1}：上一次交易的成交量除以 1 000.

5) Δt_{i-1}：从第 $i-2$ 次交易到第 $i-1$ 次交易的时间持续期.

6) BA_i：交易时刻的买卖报价差.

因为我们利用的是延迟为 1 的解释变量，所以实际的样本大小为 59 775. 对于所试用的模型 V_{i-1}，Δt_{i-1}，与 BA_i 都不是统计显著的，从而我们只运用了前三个解释变量. 采用的模型为：

192

$$\begin{aligned}\ln\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 A_{i-1} \\ \ln\left(\frac{\delta_i}{1-\delta_i}\right) &= \gamma_0 + \gamma_1 D_{i-1} \\ \ln\left(\frac{\lambda_{u,i}}{1-\lambda_{u,i}}\right) &= \theta_{u,0} + \theta_{u,1} S_{i-1} \\ \ln\left(\frac{\lambda_{d,i}}{1-\lambda_{d,i}}\right) &= \theta_{d,0} + \theta_{d,1} S_{i-1}.\end{aligned} \tag{5.30}$$

利用方程(5.29)中的对数似然函数得到的参数估计在表 5-5 中给出. 估计得到的简单模型显示了价格变化中的某些动态依赖性. 尤其是，IBM 股票的交易对交易的价格变化呈现出许多吸引人的性质：

1) 价格变化的概率依赖于上一次的价格变化. 具体地，我们有

$$P(A_i = 1 \mid A_{i-1} = 0) = 0.258, \quad P(A_i = 1 \mid A_{i-1} = 1) = 0.476.$$

这个结果显示价格变化的发生是聚类的，正如所料，大部分交易没有价格变化. 当在第$i-1$次交易中没有价格变化时，则在随后的交易中，只有大约四分之一的交易有价格变化；当在第$i-1$次交易中有价格变化时，在第 i 次交易中，价格发生变化的概率增加为大约 0.5.

2)价格变化的方向由下式控制

$$P(D_i = 1 \mid F_{i-1}, A_i) = \begin{cases} 0.483 & \text{如果 } D_{i-1} = 0(\text{即 } A_{i-1} = 0) \\ 0.085 & \text{如果 } D_{i-1} = 1, A_i = 1 \\ 0.904 & \text{如果 } D_{i-1} = -1, A_i = 1 \end{cases}$$

这个结果说明

(a)如果第 $i-1$ 次交易中没有价格变化，那么在第 i 次交易中价格上升和下降的机会是平等的；(b)连续的价格上升或者价格下降的概率是非常低的. 给定第 $i-1$ 中交易上升而且第 i 次交易中价格发生变化的条件下，第 i 次交易的价格上升的概率仅仅只有 8.6%. 然而，在给定第 $i-1$ 中交易下降而且第 i 次交易中价格发生变化的条件下，第 i 次交易的价格上升的概率大约为 90%. 因此，这个结果显示了买卖报价弹性的影响，支持了高频数据交易中的价格逆转.

3)只有很弱的证据表明大的价格变化有更大的可能性跟随另外一个大的价格变化，考虑价格增长的大小，我们有

$$S_i \mid (D_i = 1) \sim 1 + g(\lambda_{u,i}), \quad \lambda_{u,i} = 2.235 - 0.670 S_{i-1}.$$

利用几何分布的概率质量函数，我们得到，如果交易导致一个价格增长而且 $S_{i-1}=1$，那么在第 i 次交易中，价格增长一个最小变动价位的概率为 0.827. 当 $S_{i-1}=2$ 时，此概率下降为 0.709，当 $S_{i-1}=3$ 时，下降为 0.556. 因此，给定第 i 次交易中价格增长的条件下，一个大 S_i 的概率与 S_{i-1} 成反向比例.

表 5-5 方程(5.30)中对 IBM 从 1990 年 11 月 1 日到 1991 年 1 月 31 日的日内交易的 ADS 模型的参数估计

参数	β_0	β_1	γ_0	γ_1	$\theta_{u,0}$	$\theta_{u,1}$	$\theta_{d,0}$	$\theta_{d,1}$
估计	−1.057	0.962	−0.067	−2.307	2.235	−0.670	2.085	−0.509
标准差	0.104	0.044	0.023	0.056	0.029	0.050	0.187	0.139

ADS 模型与顺序概率值模型的一个区别是：ADS 模型并不要求价格变化大小中的任何截断或者分组.

5.5 持续期模型

持续期模型主要考虑交易之间的时间间隔. 较长的持续期预示着较少的交易活动，反过来表明了一个没有新消息的时期. 这样，持续期的动态行为包含了关于日内市场活动的有用信息. 利用类似于波动率的 ARCH 模型的概念，Engle 和 Russell(1998)提出了自回归条件持续期(ACD)模型来描述(大量交易的)股票时间持续期的演变. Zhang，Russell 和 Tsay(2001)扩展了 ACD 模型来解释数据中的非线性性与结构突变. 在本节中，我们引进一些简单的持续期模型. 正如前面提到的，日内交易展示了一些日模式. 因此，我们集中讨论调整的时间持续期

$$\Delta t_i^* = \Delta t_i / f(t_i) \tag{5.31}$$

其中 $f(t_i)$ 是一个确定的函数，由 Δt_i 的循环成分组成. 显然 $f(t_i)$ 依赖于标的资产(underlying

193 ~ 194

asset）以及市场的系统行为．实际应用中，有很多估计 $f(t_i)$ 的方法，但是根据统计性质没有一种方法总是优于其他方法．光滑插值是一个通常的方法，我们运用简单的二次函数与示性变量来处理日交易活动中确定的组成部分．

对于用来解释 ADS 模型所采用的 IBM 股票数据，我们假定

$$f(t_i) = \exp[d(t_i)], d(t_i) = \beta_0 + \sum_{j=1}^{7} \beta_j f_j(t_i), \tag{5.32}$$

其中

$$f_1(t_i) = -\left(\frac{t_i - 43\,200}{14\,400}\right)^2, f_3(t_i) = \begin{cases} -\left(\frac{t_i - 38\,700}{7\,500}\right)^2 & \text{若 } t_i < 43\,200 \\ 0 & \text{否则} \end{cases}$$

$$f_2(t_i) = -\left(\frac{t_i - 48\,300}{9\,300}\right)^2, f_4(t_i) = \begin{cases} -\left(\frac{t_i - 48\,600}{9\,000}\right)^2 & \text{若 } t_i \geq 43\,200 \\ 0 & \text{否则} \end{cases}$$

$f_5(t_i)$ 与 $f_6(t_i)$ 是市场开盘时第一个五分钟与第二个五分钟的示性变量（即 $f_5(.)=1$ 当且仅当 t_i 在东部时间的上午 9：30 至上午 9：35 之间）．$f_7(t_i)$ 指日交易的最后 30 分钟的示性值（即 $f_7(t_i)=1$ 当且仅当交易发生在东部时间下午 3：30 至下午 4：00）．图 5-5 画出了对 $i=1, 2, \cdots, 4$，$f_i(.)$ 的图形，其中 x 轴的时间刻度是分钟．注意到 $f_3(43\,200)=f_4(43\,200)$，其中 43 200 对应于中午 12 点．

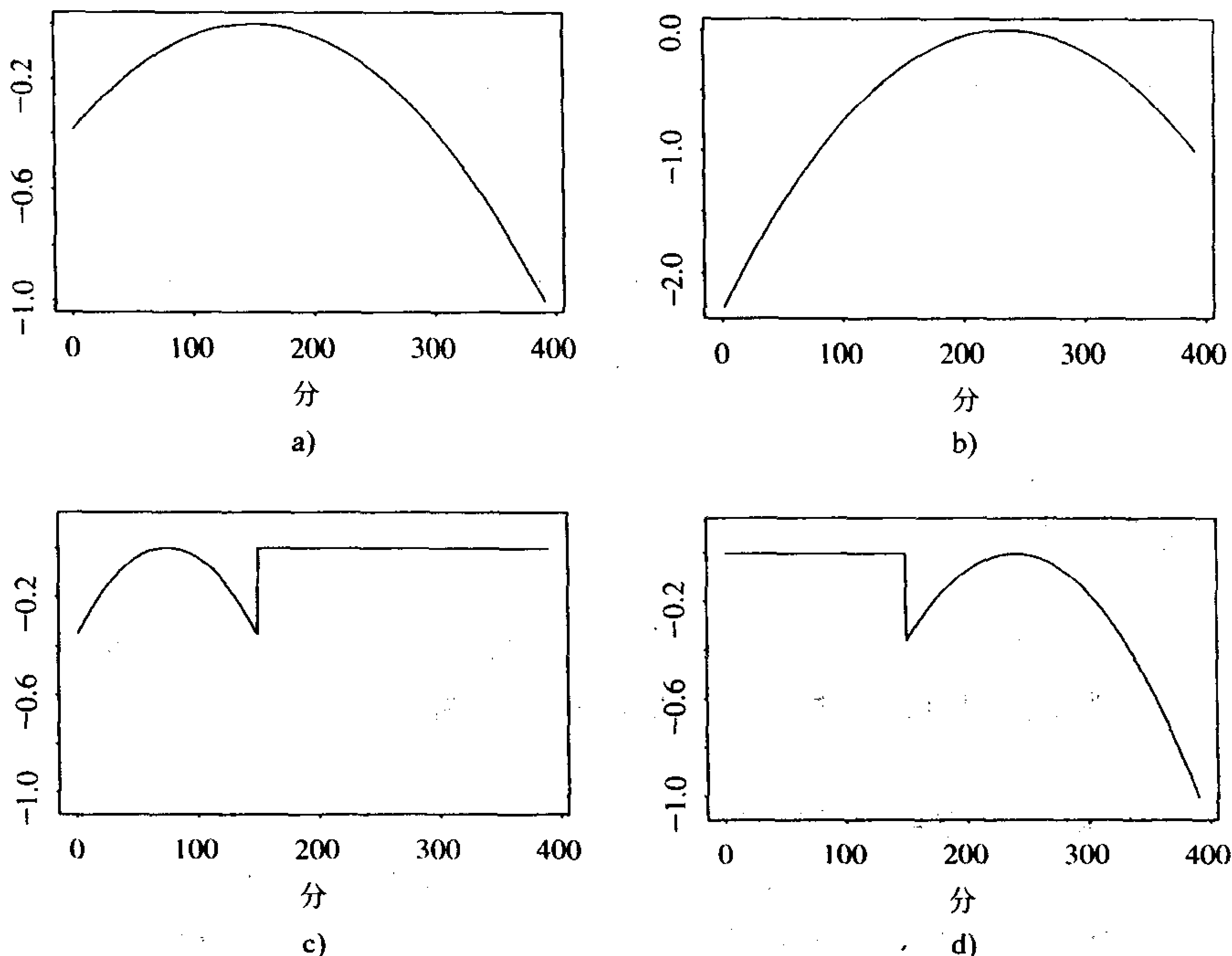


图 5-5 用来消除 IBM 日内交易持续期的确定成分的二次函数：a)～d)分别是方程(5.32)中的 $f_1(.)$ 到 $f_4(.)$

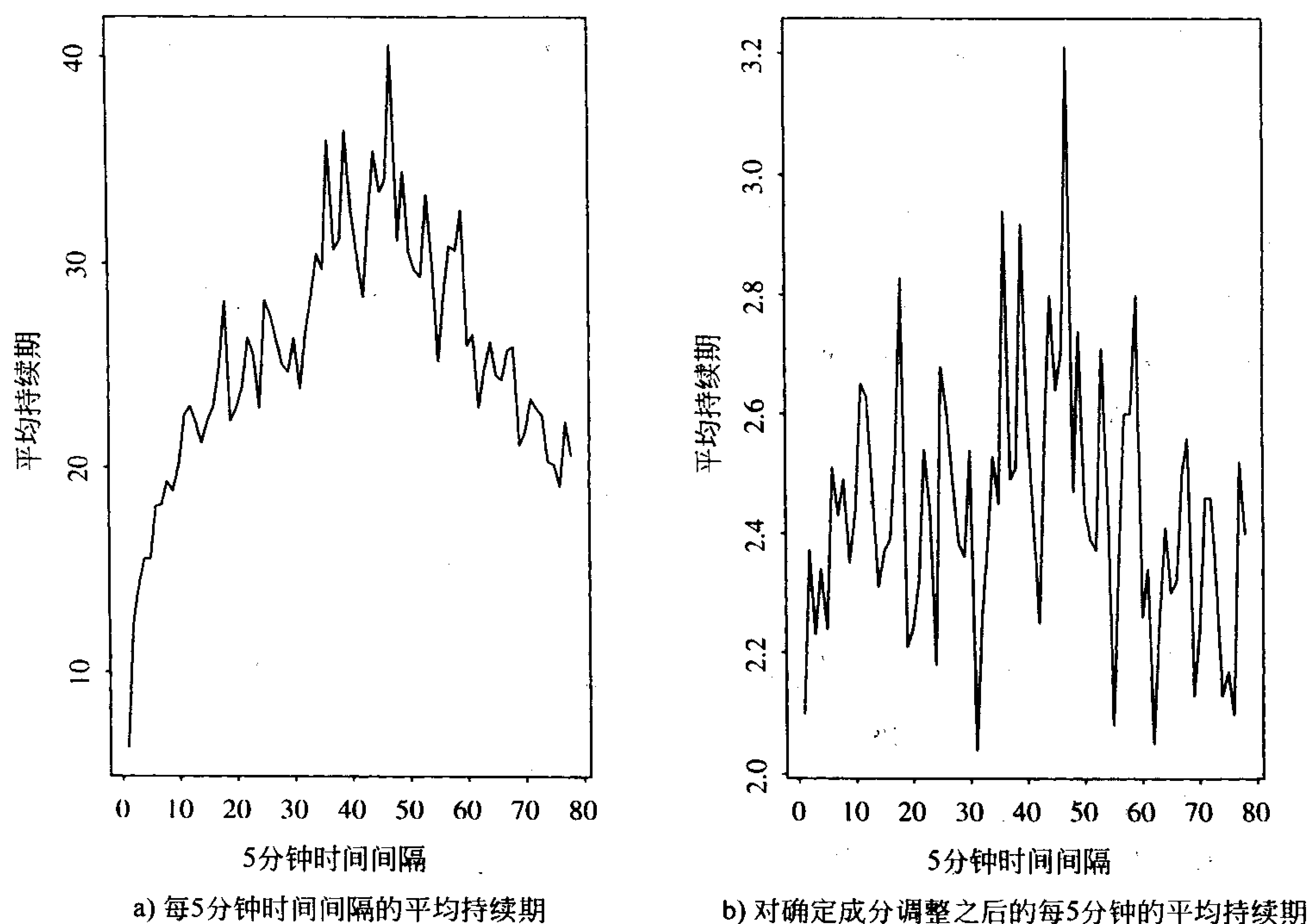
方程(5.32)中的系数 β_j 可以通过线性回归的最小二乘方法得到

$$\ln(\Delta t_i) = \beta_0 + \sum_{j=1}^{7} \beta_j f_j(t_i) + \epsilon_i.$$

拟合的模型为

$$\begin{aligned}\ln(\widehat{\Delta t_i}) = &2.555 + 0.159 f_1(t_i) + 0.270 f_2(t_i) + 0.384 f_3(t_i) \\ &+ 0.061 f_4(t_i) - 0.611 f_5(t_i) - 0.157 f_6(t_i) + 0.073 f_7(t_i).\end{aligned}$$

图 5-6 显示了 63 个交易日中每五分钟时间间隔的平均持续期的时间图，包括对于确定性部分调整前以及调整后的图形. 图 5-6a 为 Δt_i 的平均持续期，正如我们所料，它展示了一个每日模式. 图 5-6b 为 Δt_i^* 的平均持续期(也就是调整后的时间持续期)，每日模式大大削减.



a) 每5分钟时间间隔的平均持续期　　b) 对确定成分调整之后的每5分钟的平均持续期

图 5-6　IBM 从 1990 年 11 月 1 日到 1991 年 1 月 31 日的交易数据

195
~
196

5.5.1　ACD 模型

自回归条件持续期(ACD) 模型利用 GARCH 模型的思想来研究方程(5.31)中调整的时间持续期 Δt_i^* 的动态结构. 为了记号的简便，定义 $x_i = \Delta t_i^*$.

令 $\psi_i = E(x_i \mid F_{i-1})$ 表示第 $i-1$ 次交易至第 i 次交易的调整的时间持续期的条件期望，其中 F_{i-1} 为第 $i-1$ 次交易时可以得到的信息集合. 换句话说，ψ_i 为给定 F_{i-1} 的条件下期望的调

整持续期. 基本的 ACD 模型定义为

$$x_i = \psi_i \varepsilon_i \tag{5.33}$$

其中$\{\varepsilon_i\}$是独立同分布的非负随机变量序列满足$E(\varepsilon_i)=1$. 在 Engle 和 Russell(1998)中，ε_i 服从一个标准指数分布或者是标准化的韦布尔(Weibull) 分布，并且 ψ_i 假定为以下形式

$$\psi_i = \omega + \sum_{j=1}^{r} \gamma_j x_{i-j} + \sum_{j=1}^{s} \omega_j \psi_{i-j} \tag{5.34}$$

这样一个模型称为 ACD(r, s)模型. 当 ε_i 服从一个标准指数分布时，结果中的模型称为 EACD(r, s) 模型. 类似地，如果 ε_i 服从标准化的韦布尔分布，则该模型称为 WACD(r, s)模型. 有必要的话，读者可以参考附录 A，对指数分布与韦布尔分布进行一个快速的回顾.

与 GARCH 模型类似，过程 $\eta_i = x_i - \psi_i$ 是一个鞅差序列(即 $E(\eta_i \mid F_{i-1})=0$)，ACD(r, s) 模型可以写为

$$x_i = \omega + \sum_{j=1}^{\max(r,s)} (\gamma_j + \omega_j) x_{i-j} - \sum_{j=1}^{s} \omega_j \eta_{i-j} + \eta_j \tag{5.35}$$

正好是无高斯新息的 ARMA 过程的形式. 在这里，对于 $j>r$，$\gamma_j=0$；对于 $j>s$，$\omega_j=0$. 这样的表示可以用来得到 ACD 模型弱平稳性的基本条件. 例如，对方程(5.35)两边取期望，假定弱平稳性，则我们有

197

$$E(x_i) = \frac{\omega}{1 - \sum_{j=1}^{\max(r,s)} (\gamma_j + \omega_j)}.$$

因为期望持续期是正数，所以我们假定 $\omega>0$, $1 > \sum_j (\gamma_j + \omega_j)$. 作为方程(5.35)的另外一个应用，我们研究一下 EACD(1，1)模型的性质.

EACD(1，1)模型

一个 EACD(1，1)模型可以写为

$$x_i = \psi_i \varepsilon_i, \psi_i = \omega + \gamma_1 x_{i-1} + \omega_1 \psi_{i-1} \tag{5.36}$$

其中 ε_i 服从标准指数分布. 利用附录 A 中标准化指数分布的矩，我们有

$$E(\varepsilon_i) = 1, \mathrm{Var}(\varepsilon_i) = 1, E(\varepsilon_i^2) = \mathrm{Var}(\varepsilon_i) + [E(\varepsilon_i)]^2 = 2$$

假定 x_i 是弱平稳的(也就是 x_i 的头两阶矩是不随时间变化的)，那么我们导出 x_i 的方差. 首先对方程(5.36)两端取期望，我们有

$$E(x_i) = E[E(\psi_i \varepsilon_i \mid F_{i-1})] = E(\psi_i), E(\psi_i) = \omega + \gamma_1 E(x_{i-1}) + \omega_1 E(\psi_{i-1}) \tag{5.37}$$

在弱平稳性的条件下，$E(\psi_i)=E(\psi_{i-1})$，因此方程(5.37)给出

$$\mu_x \equiv E(x_i) = E(\psi_i) = \frac{\omega}{1-\gamma_1-\omega_1} \tag{5.38}$$

又因为 $E(\varepsilon_i^2)=2$，我们有

$$E(x_i^2) = E[E(\psi_i^2\varepsilon_i^2 \mid F_{i-1})] = 2E(\psi_i^2)$$

在方程(5.36)中对 ψ_i 的平方取期望，并且利用 ψ_i 和 x_i 的弱平稳性质，通过某些代数运算，我们得到

$$E(\psi_i^2) = \mu_x^2 \times \frac{1-(\gamma_1+\omega_1)^2}{1-2\gamma_1^2-\omega_1^2-2\gamma_1\omega_1} \tag{5.39}$$

最后，利用 $\mathrm{Var}(x_i)=E(x_i^2)-[E(x_i)]^2$，$E(x_i^2)=2E(\psi_i^2)$，我们有

$$\mathrm{Var}(x_i) = 2E(\psi_i^2)-\mu_x^2 = \mu_x^2 \times \frac{1-\omega_1^2-2\gamma_1\omega_1}{1-\omega_1^2-2\gamma_1\omega_1-2\gamma_1^2},$$

其中 μ_x 如方程(5.38)中定义. 这个结果显示，为了得到不随时间变化的无条件方差，方程(5.36)中的 EACD(1，1)模型必须满足 $1>2\gamma_1^2+\omega_1^2+2\gamma_1\omega_1$. WACD(1，1)模型的方差可以利用同样的方法以及标准化韦布尔分布的前两阶矩得到.

带有广义伽马分布的 ACD 模型

在统计文献中，强度函数经常根据危险率函数来表示. 正如附录 B 中所示，EACD 模型的危险率函数是不随时间变的常数，WACD 模型的危险率函数是一个单调函数. 这些危险率函数在实际应用中相当受限制，因为股票交易的强度函数可能不是固定的，或者不是随时间单调变化的. 为了增加相应的危险率函数的灵活性，Zhang，Russell 和 Tsay(2001) 对 ε_i 采用了一个(标准化的)广义伽马(Gamma)分布. 对于广义伽马分布的基本性质可以参见附录 A. 结果中的危险率函数可以假定不同的模式，包括 U 型或者倒转的 U 型. 我们将新息服从广义伽马分布的 ACD 模型称为 GACD(r，s)模型. 198

5.5.2 模拟

为了解释 ACD 过程，我们从 ACD(1，1)模型

$$x_i = \psi_i\varepsilon_i, \psi_i = 0.3+0.2x_{i-1}+0.7\psi_{i-1} \tag{5.40}$$

利用 ε_i 的两种不同的新息分布，各产生了 500 个观测值. 在第一种情形，假定 ε_i 服从参数 $\alpha=1.5$ 的标准化的韦布尔分布；在第二种情形，假定 ε_i 服从参数 $\kappa=1.5$，$\alpha=0.5$ 的一个(标准化的)广义伽马分布.

图 5-7a 显示了 WACD(1，1)序列的时间图，图 5-8a 为 GACD(1，1)序列的时间图. 图 5-9画出了两个模拟序列的直方图.

两个模型的区别是很显然的. 最后，在图 5-10a 与图 5-11b 中，分别画出了两个模拟序列的样本 ACF，从中可以清楚地看到序列的相依性.

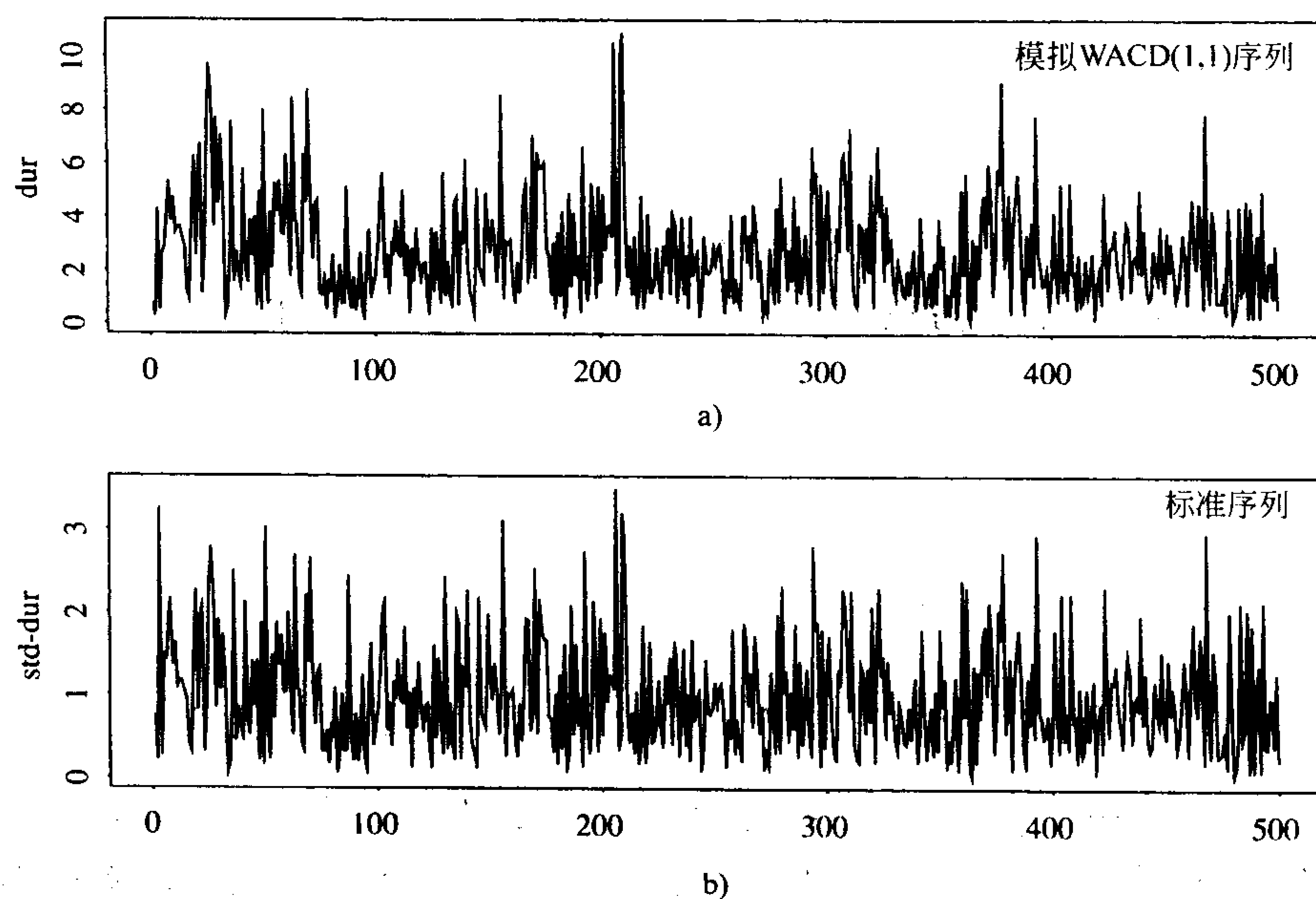


图 5-7 方程(5.40)模拟的 WACD(1, 1)序列：a)原始序列，b)估计后的标准化序列．共有 500 个观测值

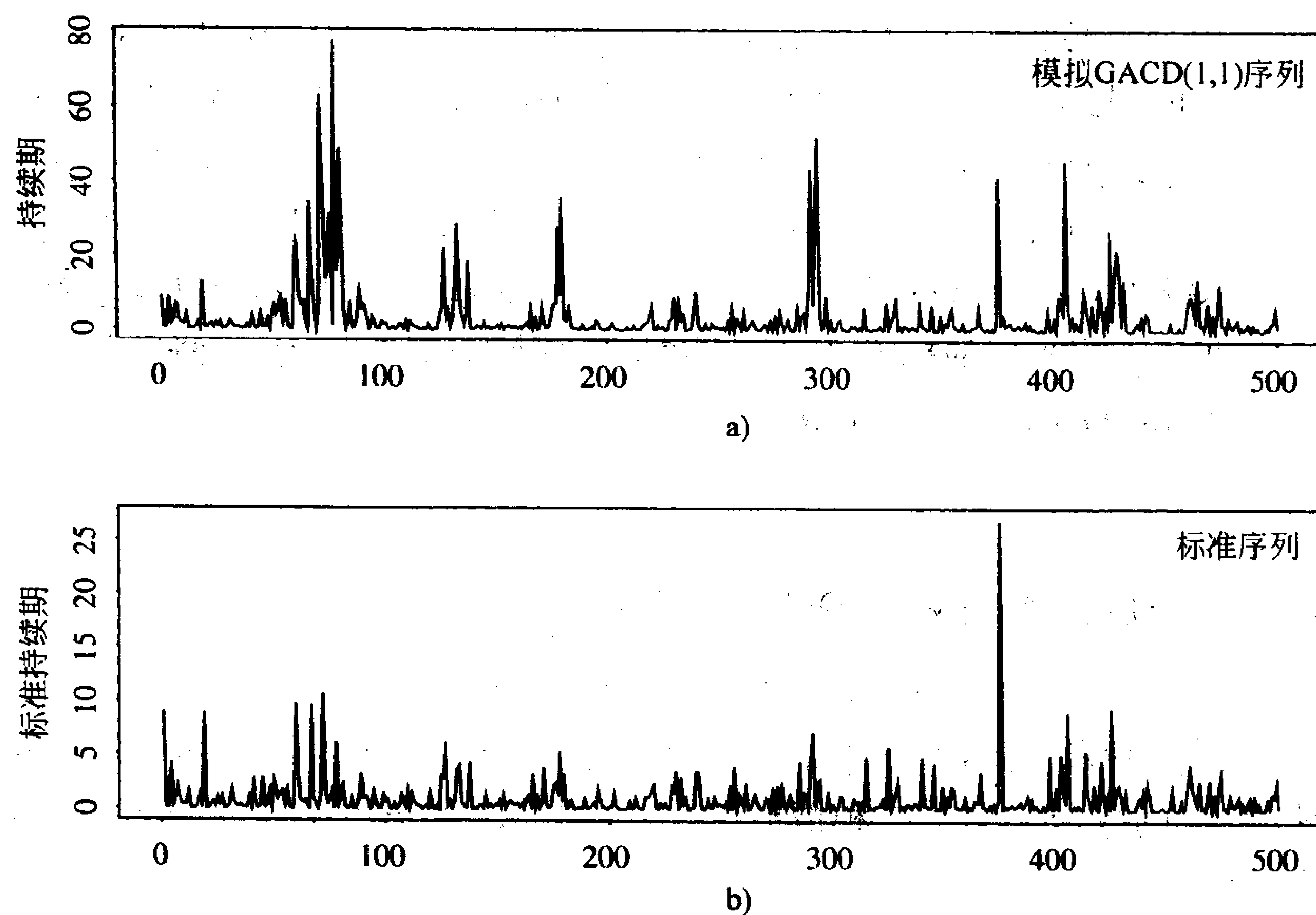


图 5-8 方程(5.40)模拟的 GACD(1, 1)序列：a)原始序列，b)估计后的标准化序列．共有 500 个观测值

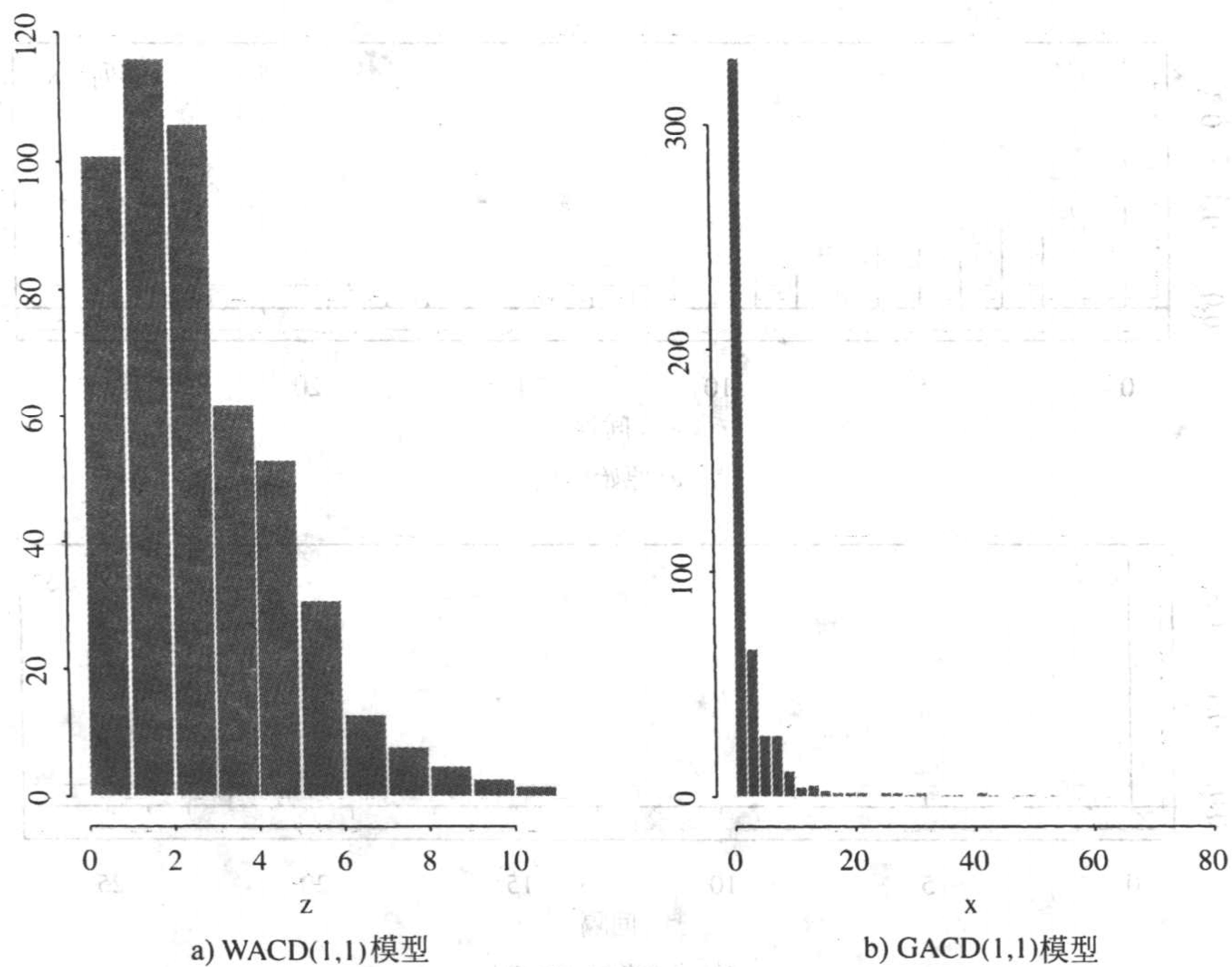


图 5-9 具有 500 个观测值的模拟持续期过程的直方图

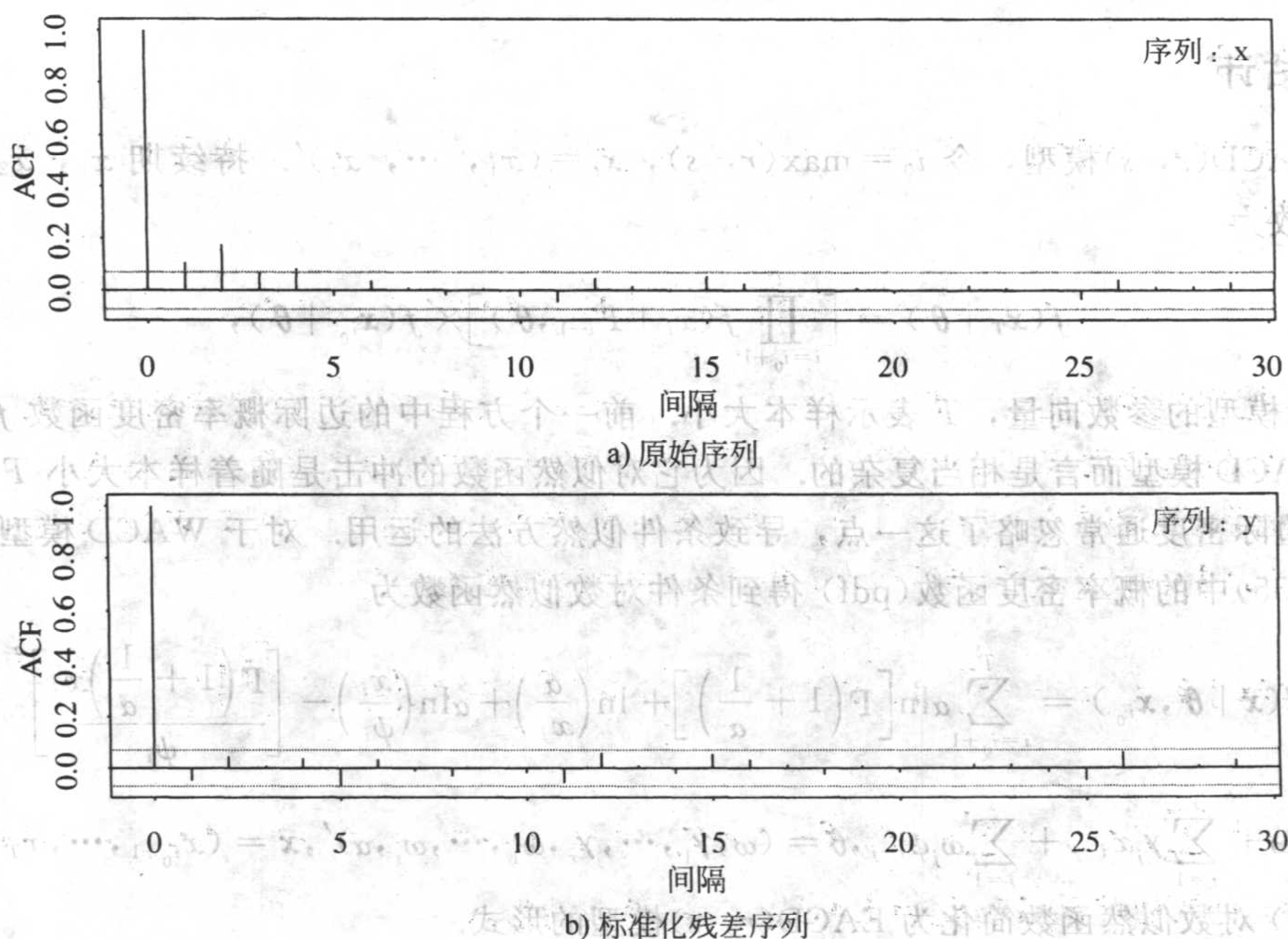
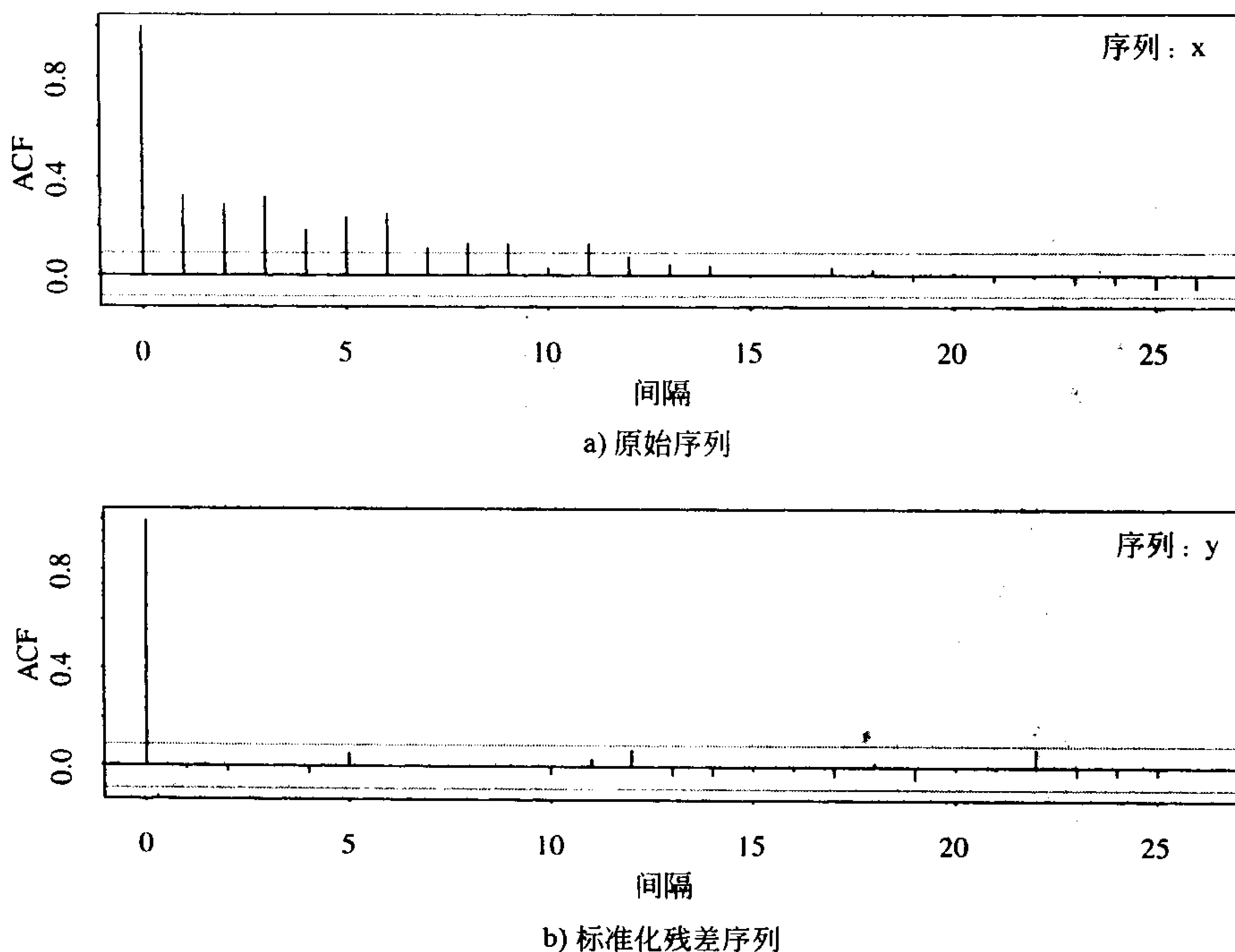


图 5-10 具有 500 个观测值的模拟 WACD(1，1)序列的样本自相关函数



a) 原始序列

b) 标准化残差序列

图 5-11 具有500个观测值的模拟GACD(1，1)序列的样本自相关函数

5.5.3 估计

对于ACD(r，s)模型，令 $i_0=\max(r, s)$，$\boldsymbol{x}_t=(x_1, \cdots, x_t)'$. 持续期 $x_1, x_2, \cdots, x_T$ 的似然函数为

$$f(\boldsymbol{x}_T \mid \boldsymbol{\theta})=\left[\prod_{i=i_0+1}^{T} f(x_i \mid F_{i-1}, \boldsymbol{\theta})\right] \times f(\boldsymbol{x}_{i_0} \mid \boldsymbol{\theta}),$$

其中$\boldsymbol{\theta}$表示模型的参数向量，T表示样本大小. 前一个方程中的边际概率密度函数 $f(\boldsymbol{x}_{i_0} \mid \boldsymbol{\theta})$ 对广义的ACD模型而言是相当复杂的. 因为它对似然函数的冲击是随着样本大小 T 的增加而递减的，边际密度通常忽略了这一点，导致条件似然方法的运用. 对于WACD模型，我们利用方程(5.55)中的概率密度函数(pdf) 得到条件对数似然函数为

$$l(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{x}_{i_0})=\sum_{i=i_0+1}^{T} \alpha \ln\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]+\ln\left(\frac{\alpha}{x_i}\right)+\alpha \ln\left(\frac{x_i}{\psi_i}\right)-\left[\frac{\Gamma\left(1+\frac{1}{\alpha}\right) x_i}{\psi_i}\right]^{\alpha} \quad (5.41)$$

其中 $\psi_i=\omega+\sum_{j=1}^{r} \gamma_j x_{i-j}+\sum_{j=1}^{s} \omega_j \psi_{i-j}$, $\boldsymbol{\theta}=(\omega, \gamma_1, \cdots, \gamma_r, \omega_1, \cdots, \omega_s, \alpha)'$, $\boldsymbol{x}=(x_{i_0+1}, \cdots, x_T)'$. 当 $\alpha=1$ 时,(条件) 对数似然函数简化为EACD(r，s)模型的形式.

对一个GACD(r，s)模型，条件对数似然函数为

$$l(x \mid \theta, x_{i_0}) = \sum_{i=i_0+1}^{T} \ln\left(\frac{\alpha}{\Gamma(\kappa)}\right) + (\kappa\alpha - 1)\ln(x_i) - \kappa\alpha\ln(\lambda\psi_i) - \left(\frac{x_i}{\lambda\psi_i}\right)^{\alpha} \qquad (5.42)$$

其中$\lambda=\Gamma(\kappa)/\Gamma\left(\kappa+\frac{1}{\alpha}\right)$，参数向量$\theta$也包含$\kappa$. 如我们料想的，当$\kappa=1$，$\lambda=1/\Gamma\left(1+\frac{1}{\alpha}\right)$时，方程(5.42)中的对数似然函数简化为方程(5.41)中模型WACD(r, s)的似然函数形式. 这个对数似然函数可以改写为许多形式来简化估计.

199 ~ 202

在一些正则条件下，条件最大似然估计是渐近正态的，详见Engle和Russell(1998)及其参考文献. 实际中，一旦指定了持续期模型，对于感兴趣的问题可以用模拟的方法得到有限样本的参考分布.

例5.3 (模拟的ACD(1, 1)序列的续)考虑方程(5.40)中模拟的WACD(1, 1)以及GACD(1, 1)序列，利用条件似然方法得到的结果列于表5-6中，估计看上去是合理的. 令$\hat{\psi}$表示ψ_i的1步向前预测，$\hat{\epsilon}_i=x_i/\hat{\psi}_i$为标准化的序列，可以认为是序列标准化的残差. 如果模型指定地充分，$\{\hat{\epsilon}_i\}$应该以独立同分布随机变量序列的形式运动. 图5-7b与图5-8b显示了两个模型$\hat{\epsilon}_i$的时间图. 拟合的两个模型的$\hat{\epsilon}_i$的样本ACF分别在图5-10b与图5-11b中给出. 很显然，在$\hat{\epsilon}_i$序列中没有发现显著的序列相关.

表5-6 对模拟的ACD(1, 1)模型的500个观测值序列的估计结果：(a)WACD(1, 1)序列(b)GACD(1, 1)序列

(a)WACD(1, 1)模型					
参数	ω	γ_1	ω_1	α	
真实值	0.3	0.2	0.7	1.5	
估计值	0.364	0.100	0.767	1.477	
标准误差	(0.139)	(0.025)	(0.060)	(0.052)	
(b)GACD(1, 1)模型					
参数	ω	γ_1	ω_1	α	κ
真实值	0.3	0.2	0.7	0.5	1.5
估计值	0.401	0.343	0.561	0.436	2.077
标准误差	(0.117)	(0.074)	(0.065)	(0.078)	(0.653)

例5.4 作为持续期模型的说明，我们考虑IBM股票从1990年11月1日至1990年11月7日连续五个交易日内的交易间的持续期. 集中于正的交易持续期，我们共有3 534个观测值. 另外，将数据作了调整，消除了方程(5.32)中的确定成分. 也就是，我们采用了方程(5.31)中定义的3 534个正的调整持续期.

图5-12a显示了1990年11月的前五个交易日内调整的(正的)持续期的时间图，图5-13a给出了序列的样本ACF. 在调整的持续期中，还存在一些序列相关性. 我们对数据拟合一个WACD(1, 1)模型，得到的模型为

$$x_i = \psi_i\epsilon_i, \psi_i = 0.169 + 0.064x_{i-1} + 0.885\psi_{i-1} \qquad (5.43)$$

203

其中$\{\epsilon_i\}$是一个独立同分布的随机变量序列，服从参数$\hat{\alpha}=0.879(0.012)$的标准化的韦布尔分

布，这里 0.012 为估计的标准差．方程(5.43)中估计的标准差分别为 0.039，0.010 和 0.018．所有估计的 t 比率都大于 4.2，说明估计在 1%水平下是显著的．图 5-12b 给出了 $\hat{\varepsilon}_i = x_i/\hat{\psi}_i$ 的时间图，图 5-13b 提供了 $\hat{\varepsilon}_i$ 的样本 ACF．$\hat{\varepsilon}_i$ 序列的 Ljung-Box 统计量(以下简称为 L-B 统计量——译者注)为 $Q(10)=4.96$，$Q(20)=10.75$．显然，标准化的新息没有显著的序列相关性．事实上，平方序列 $\{\hat{\varepsilon}_i^2\}$ 的样本自相关性也很小，其 $Q(10)=6.20$，$Q(20)=11.16$，进一步确认了标准化的新息之间缺少序列相依性．另外，参数为 $\alpha=0.879$ 的标准化的韦布尔分布的均值和标准差分别为 1.00 和 1.14，这些数字也接近于 $\{\hat{\varepsilon}_i\}$ 的样本均值 1.01 和标准差 1.22．拟合的模型看上去是合适的．

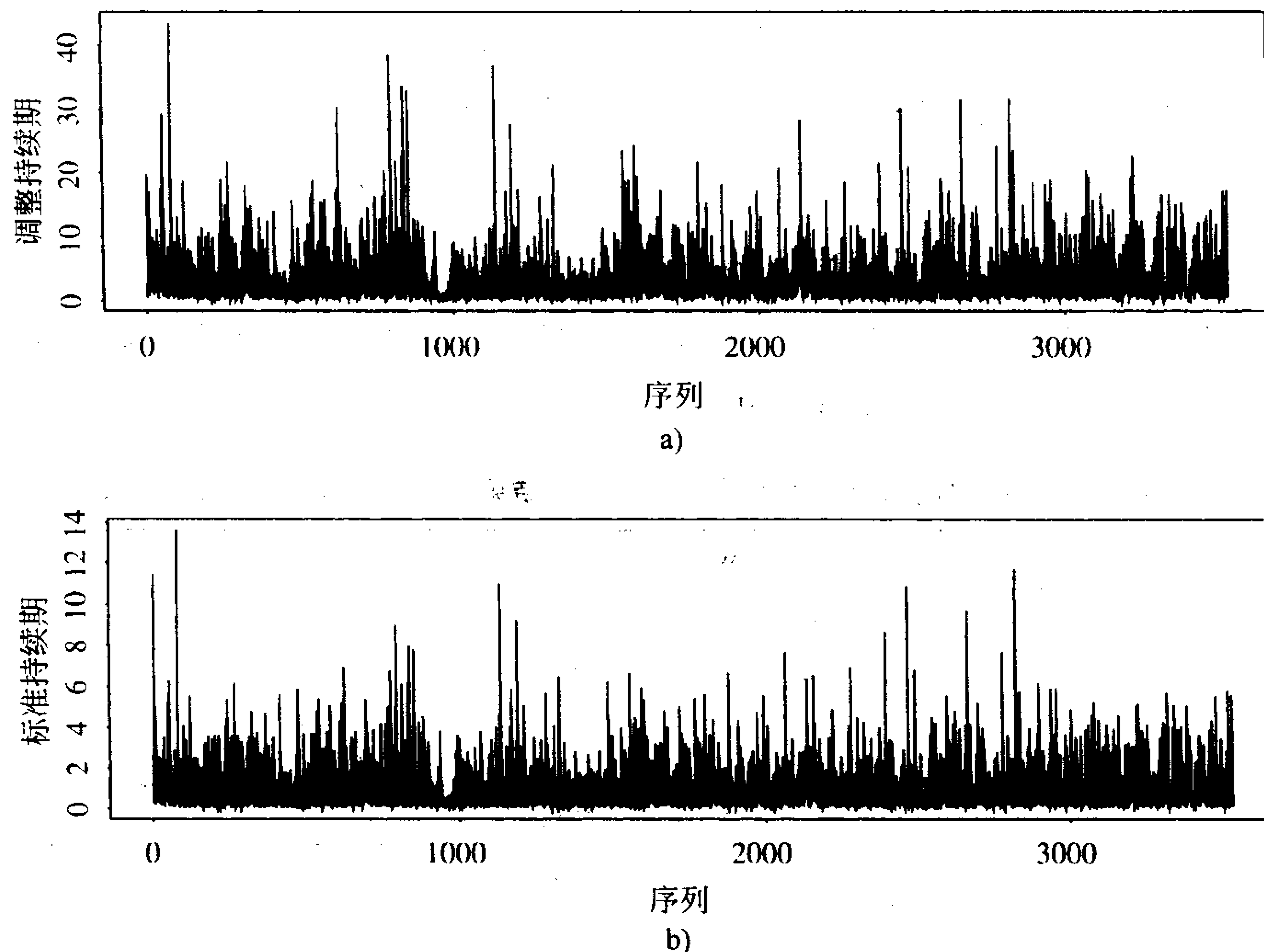


图 5-12 在 1990 年 11 月的前 5 个交易日交易的 IBM 股票的持续期的时间图：a)调整的序列，b)WACD(1，1)模型的标准化新息．有 3 534 个非零的持续期数据

在模型(5.43)中，估计系数 $\hat{\gamma}_1+\hat{\omega}_1 \approx 0.949$，隐含了调整持续期的某种持续性．调整持续期的期望为 $0.169/(1-0.064-0.885)=3.31$ 秒，这接近于调整持续期的样本均值 3.29．标准化韦布尔分布的形状参数的估计 $\hat{\alpha}$ 为 0.879，小于 1 但是接近于 1．这样，条件危险率函数是以缓慢的速率单调递减的．

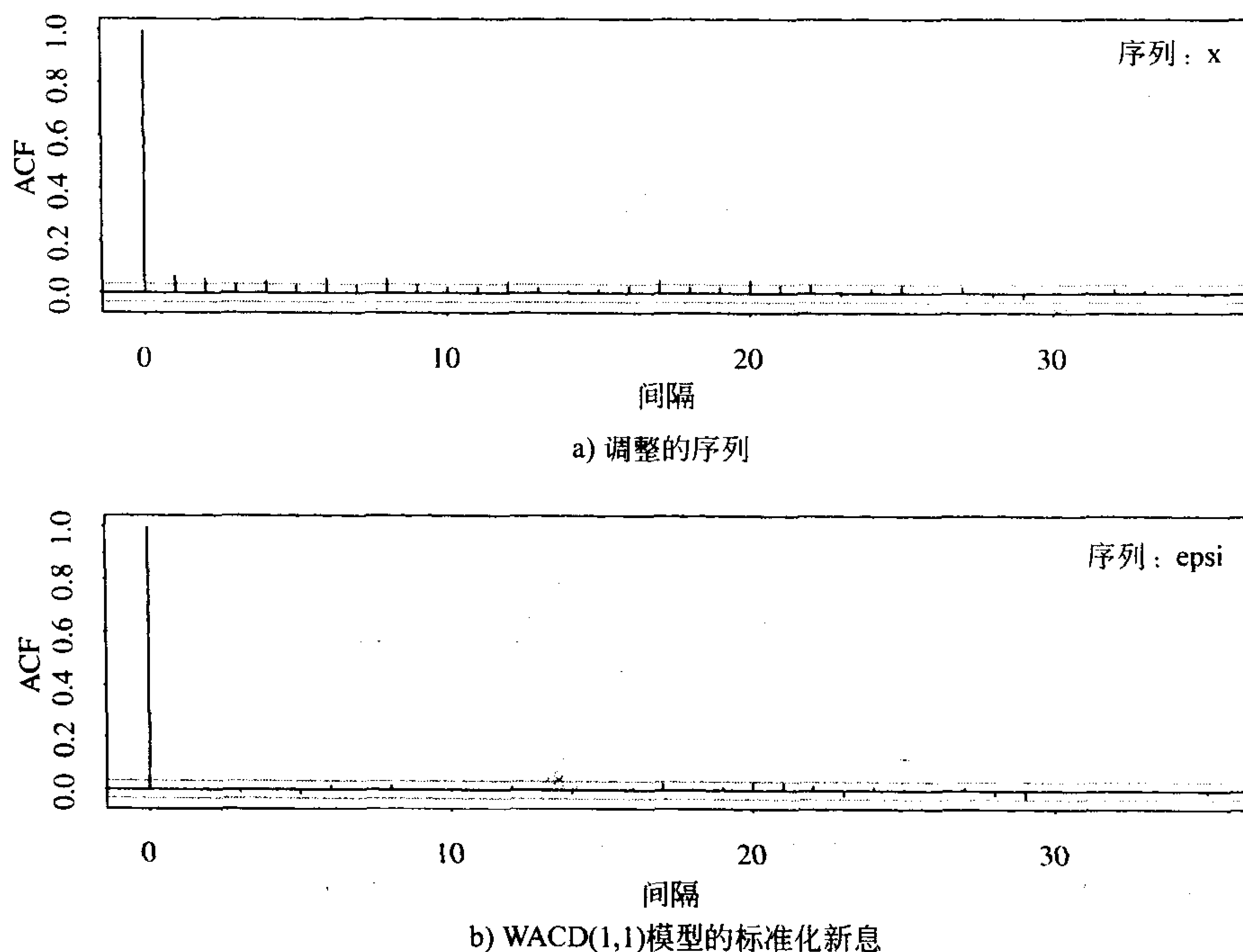
[204]

如果对新息采用广义伽马分布函数，那么拟合的 GACD(1，1)模型为

$$x_i = \psi_i \varepsilon_i, \psi_i = 0.141 + 0.063 x_{i-1} + 0.897 \psi_{i-1} \tag{5.44}$$

其中 $\{\varepsilon_i\}$ 服从方程(5.56)的标准化的广义伽马分布，参数 $\kappa=4.248(1.046)$，参数 $\alpha=0.395(0.053)$，括号内的数字表示估计的标准误差．方程(5.44)中三个参数的标准误差分别为 0.041，0.010 和 0.019，所有的估计在 1%水平下都是统计上显著的，标准化的新息过程 $\{\hat{\varepsilon}_i\}$ 及其平方序列都没有显

著的序列相关，这里$\hat{\varepsilon}_i = x_i/\hat{\psi}_i$是根据模型(5.44)定义的. 具体地，对$\hat{\varepsilon}_i$过程，我们有$Q(10)=4.95$，$Q(20)=10.28$；对$\hat{\varepsilon}_i^2$序列，我们有$Q(10)=6.36$，$Q(20)=10.89$.



a) 调整的序列

b) WACD(1,1)模型的标准化新息

图 5-13 在 1990 年 11 月的前 5 个交易日交易的 IBM 股票的调整持续期的样本自相关函数

模型(5.44)的期望持续期是 3.52，稍高于方程(5.43)中 WACD(1，1)模型. 类似地，模型(5.44)中的持续参数$\hat{\gamma}_1+\hat{\omega}_1$也稍高于 0.96.

205

注释： EACD 模型的估计可以利用经微小修改的 ARCH 模型的程序来实现，可以参见 Engle 和 Russell(1998). 在本书中，我们用作者发展的 RATS 程序或者一些 Fortran 程序来估计持续期模型. 有限的经验告诉我们估计一个 GACD 模型比估计一个 EACD 模型或 WACD 模型都更加困难. 用来估计 WACD 与 GACD 模型的 RATS 程序在附录 C 中给出.

5.6 非线性持续期模型

高频数据中也通常会发现非线性性质，作为说明，我们对例 5.4 中对 IBM 交易的持续期拟合的 WACD(1，1)模型的标准化新息$\hat{\varepsilon}_i$利用第 4 章中讨论的一些非线性检验，参见方程(5.43). 根据 AR(4)模型，在表 5-7 中的(a)部分给出检验结果. 正如从例 5.4 的模型诊断中料到的，Ori-F 检验显示标准化新息中没有二次非线性. 然而，TAR-F 检验统计量却显示了很强的非线性.

根据表5-7的检验结果，我们对IBM的日内持续期考虑一个2体制的门限持续期模型，门限变量为x_{t-1}(即1步延迟调整持续期). 门限值的估计为3.79，拟合的门限WACD(1，1)模型是$x_i=\psi_i\epsilon_i$，其中

$$\psi_i=\begin{cases}0.020+0.257x_{i-1}+0.847\psi_{i-1}, & \epsilon_i\sim w(0.901) \quad 若\ x_{i-1}\leqslant 3.79\\ 1.808+0.027x_{i-1}+0.501\psi_{i-1}, & \epsilon_i\sim w(0.845) \quad 若\ x_{i-1}>3.79\end{cases} \tag{5.45}$$

这里$w(\alpha)$表示参数α的标准化韦布尔分布. 两种体制下观测值的数量分别为2 503和1 030. 在方程(5.45)中，第一个体制中参数的标准误差分别为0.043，0.041，0.024和0.014；而第二个体制中参数的标准差分别为0.526，0.020，0.147和0.020.

表 5-7

(a)WACD(1，1)模型的正态化新息					
类型	Ori-F	Tar-F(1)	Tar-F(2)	Tar-F(3)	Tar-F(4)
检验	0.343	3.288	3.142	3.128	0.297
p值	0.969	0.006	0.008	0.008	0.915
(b)门限WACD(1，1)模型的正态化新息					
类型	Ori-F	Tar-F(1)	Tar-F(2)	Tar-F(3)	Tar-F(4)
检验	0.163	0.746	1.899	1.752	0.270
p值	0.998	0.589	0.091	0.119	0.929

注：对IBM从1990年11月1日至11月7日的交易持续期的非线性检验，只利用了日内持续期. Tar-F检验括号内的数表示时间延迟.

考虑方程(5.45)表示的门限WACD(1，1)模型的标准化新息$\hat{\epsilon}_i=x_i/\hat{\psi}_i$，得到$\hat{\epsilon}_i$的$Q(12)=9.8$，$Q(24)=23.9$；$\hat{\epsilon}_i^2$的$Q(12)=8.0$，$Q(24)=16.7$. 这样，$\hat{\epsilon}_i$序列与$\hat{\epsilon}_i^2$序列均无显著的序列相关性. 而且，对这个新的标准化新息序列$\hat{\epsilon}_i$，应用前面同样的非线性检验，没有检验出非线性，结果见表5-7中(b)部分. 因此，方程(5.45)中的2体制WACD(1，1)模型是合适的.

如果我们将两体制分为大量交易与很少交易两个时期，那么对IBM股票门限模型显示，由日内交易持续期度量的交易动态性在大量交易与较少交易两个时期是不同的(即使在日内模式调整之后). 这并不令人惊奇，因为市场活动经常是新的信息和其他信息的到达驱动的.

方程(5.45)估计出的门限WACD(1，1)模型包含一些不显著的参数，我们将模型提炼得到如下结果

$$\psi_i=\begin{cases}0.225x_{i-1}+0.867\psi_{i-1}, & \epsilon_i\sim w(0.902) \quad 若\ x_{i-1}\leqslant 3.79\\ 1.618+0.614\psi_{i-1}, & \epsilon_i\sim w(0.846) \quad 若\ x_{i-1}>3.79\end{cases}$$

提炼出的全部模型的参数估计都是高度显著的，标准化的新息$\hat{\epsilon}_i=x_i/\hat{\psi}_i$的L-B统计量显示$Q(10)=5.91(0.82)$，$Q(20)=16.04(0.71)$；$\hat{\epsilon}_i^2$的L-B统计量显示$Q(10)=5.35(0.87)$，$Q(20)=15.20(0.76)$，这里括号内的数字是$p$值. 因此，提炼出的模型是合适的. 用来估计前述模型的RATS程序在附录C中给出.

5.7 价格变化和持续期的二元模型

本节中，我们引进一个考虑价格变化和相应持续期的联合过程的模型．如前所述，许多股票的日内交易是无价格变化的，那些交易与交易强度高度相关，但是它们不包含价格运动的直接的信息．因此，为了简化价格变化建模中的复杂性，我们集中讨论导致价格变化的交易，并考虑同时描述价格变化和相应时间持续期(PCD)的多元动态性模型．

继续利用上一节中同样的记号，但是定义改为价格变化的交易．令 t_i 是资产在第 i 次价格变化的日历时间．同前面一样，t_i 是从一个交易日的午夜开始以秒为单位测量的．令 P_{t_i} 表示第 i 次价格发生变化时的交易价格，$\Delta t_i=t_i-t_{i-1}$ 是价格变化的时间持续期．另外，令 N_i 表示时间间隔(t_{i-1}, t_i)中无价格变化时的交易数量．这个新变量用来表示无价格变化期间的交易强度．最后，令 D_i 表示第 i 次价格变化的方向，当价格上升时，$D_i=1$，当价格下降时，$D_i=-1$．令 S_i 表示以最小价位变动单位的第 i 次价格变化的大小．在新的定义下，股票价格随时间的变化为

206～207

$$P_{t_i}=P_{t_{i-1}}+D_iS_i \tag{5.46}$$

对第 i 次价格变化的交易数据包括$\{\Delta t_i, N_i, D_i, S_i\}$，PCD 模型关心的是对$(\Delta t_i, N_i, D_i, S_i)$的联合分析．

注释：集中于与价格变化相联系的交易可以大大地降低样本的大小．例如，考虑 IBM 股票从 1990 年 11 月 1 日至 1991 年 1 月 31 日的日内数据，共有 60 265 个日内交易，但是它们中只有 19 022 个引起了价格变化．另外，价格变化之间的时间持续期中没有日内模式．

为了解释所有交易的价格运动与只跟价格变化相关的交易的价格运动之间的关系，我们考虑 IBM 股票在 1990 年 11 月 21 日的日内交易，在这一天的正常交易时间有 726 个交易，只有 195 个交易引起了价格变化．图 5-14 显示了两种情况下价格序列的时间图．正如我们想像的，价格序列是相同的．

给定 F_{i-1} 的条件下，PCD 模型将$(\Delta t_i, N_i, D_i, S_i)$的联合分布分解为

$$\begin{aligned}&f(\Delta t_i,N_i,D_i,S_i\mid F_{i-1})\\&=f(S_i\mid D_i,N_i,\Delta t_i,F_{i-1})f(D_i\mid N_i,\Delta t_i,F_{i-1})f(N_i\mid \Delta t_i,F_{i-1})f(\Delta t_i\mid F_{i-1})\end{aligned} \tag{5.47}$$

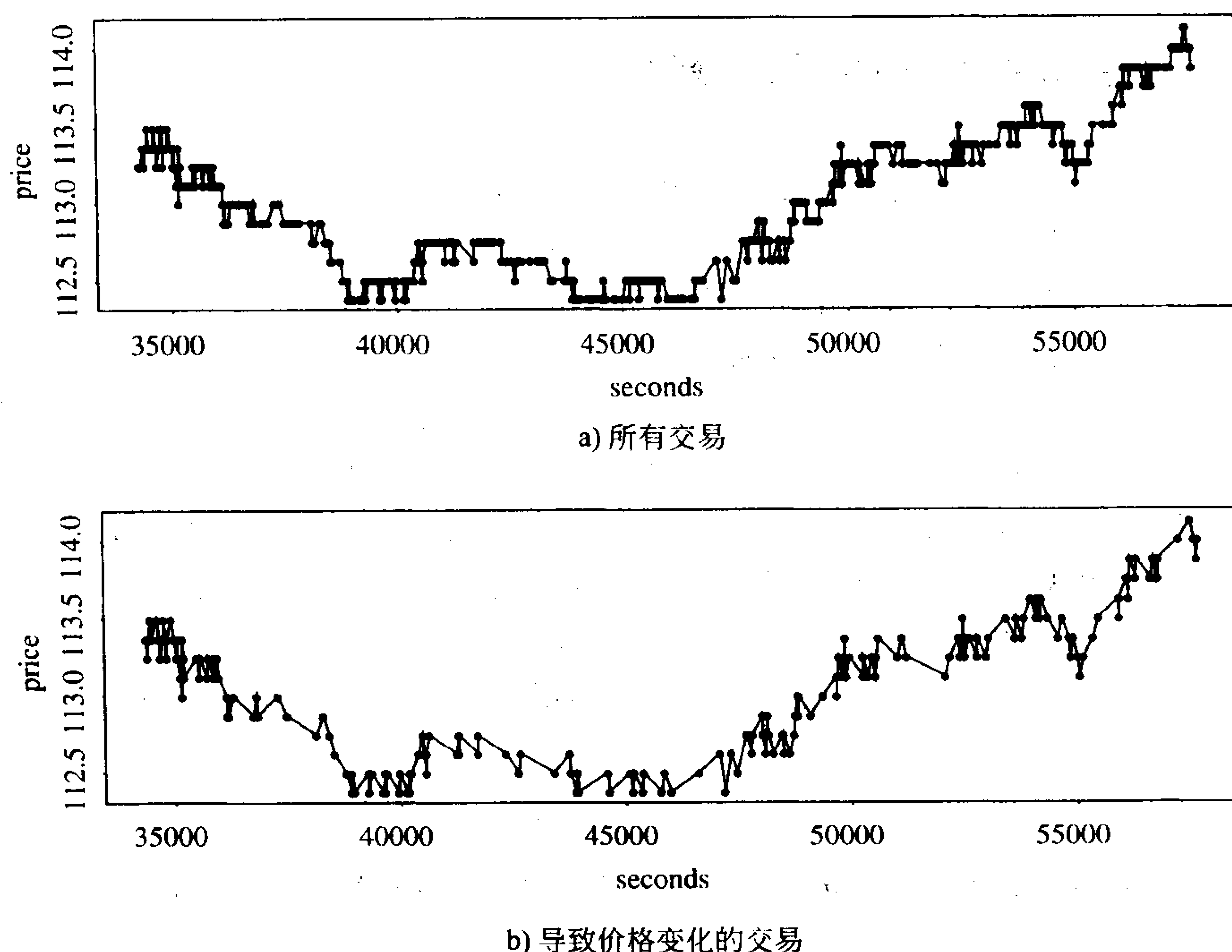
这个分解使得我们能够对条件分布指定合适的经济计量模型，从而简化建模．有许多方式对条件分布指定模型，一个恰当的指定可能依赖于所研究的资产．这里我们采用 McCulloch 和 Tsay(2000)使用的指定，对离散取值的变量利用广义的线性模型，对连续取值的变量 $\ln(\Delta t_i)$ 利用了一个时间序列模型．

对价格变化之间的时间持续期，我们采用的模型为

$$\ln(\Delta t_i)=\beta_0+\beta_1\ln(\Delta t_{i-1})+\beta_2S_{i-1}+\sigma\epsilon_i \tag{5.48}$$

其中 σ 是正数，$\{\epsilon_i\}$是 iid$N(0, 1)$的随机变量序列．这是关于延迟变量的多元线性回归模型．如果有必要，可以加入其他的解释变量．为了确保时间持续期为正，我们利用了对数变换．

208



a) 所有交易

b) 导致价格变化的交易

图 5-14 IBM 股票在 1990 年 11 月 21 日的日内交易价格的时间图

N_i 的条件模型又分解为两部分，因为经验数据显示 N_i 集中在 0 点. N_i 模型的第一部分为 logit 模型

$$p(N_i = 0 \mid \Delta t_i, F_{i-1}) = \text{logit}[\alpha_0 + \alpha_1 \ln(\Delta t_i)] \tag{5.49}$$

其中 $\text{logit}(x)=\exp(x)/[1+\exp(x)]$，而模型的第二部分为

$$N_i \mid (N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \lambda_i = \frac{\exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]}{1 + \exp[\gamma_0 + \gamma_1 \ln(\Delta t_i)]} \tag{5.50}$$

其中～意思是“服从的分布为”，$g(\lambda)$ 表示参数为 λ 的几何分布，λ 在区间(0，1)上取值.

方向 D_i 的模型为

$$D_i \mid (N_i, \Delta t_i, F_{i-1}) = \text{sign}(\mu_i + \sigma_i \epsilon) \tag{5.51}$$

209 其中 ϵ 是 $N(0,1)$ 的随机变量，而且

$$\mu_i = \omega_0 + \omega_1 D_{i-1} + \omega_2 \ln(\Delta t_i),$$

$$\ln(\sigma_i) = \beta \left| \sum_{j=1}^{4} D_{i-j} \right| = \beta |D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4}|.$$

换句话说，D_i 是由均值为 μ_i、方差为 σ_i^2 的正态随机变量控制的. 上述模型的特殊特征是函数

$\ln(\sigma_i)$. 对日内交易，一个关键的性质是两个连续价格变化之间的价格逆转. 这个性质是通过在均值方程中 D_i 对 D_{i-1} 的依赖允许 ω_1 为负值来建模的. 然而，价格运动中存在很少的局部趋势. 当过去数据呈现出局部趋势时，上述的方差方程通过增加价格运动方向的不确定性，允许这样的局部趋势存在于模型中. 对于一个固定均值的正态分布，增加方差使得一个随机移动有同样的机会取正值或负值. 这又增加了序列全部为正的移动与全部为负的移动的机会. 这样一个序列产生了价格运动的局部趋势.

为了允许正负价格运动之间的不同动态性，我们对价格变化的大小运用不同的模型，具体地，我们有

$$S_i \mid (D_i = -1, N_i, \Delta t_i, F_{i-1}) \sim p(\lambda_{d,i}) + 1, \tag{5.52}$$

$$\ln(\lambda_{d,i}) = \eta_{d,0} + \eta_{d,1} N_i + \eta_{d,2} \ln(\Delta t_i) + \eta_{d,3} S_{i-1}$$

$$S_i \mid (D_i = 1, N_i, \Delta t_i, F_{i-1}) \sim p(\lambda_{u,i}) + 1, \tag{5.53}$$

$$\ln(\lambda_{u,i}) = \eta_{u,0} + \eta_{u,1} N_i + \eta_{u,2} \ln(\Delta t_i) + \eta_{u,3} S_{i-1}$$

其中 $p(\lambda)$表示参数为 λ 的泊松(Poisson)分布，在大小中加入 1 是因为当价格变化时，最小的大小为一个最小变动价位.

方程(5.48)～(5.53)中指定的模型可以通过最大似然方法或 MCMC 方法来联合估计. 根据方程(5.47)，这些模型包含了六个可以单独估计的条件模型.

例 5.5 考虑 IBM 股票在 1990 年 11 月 21 日的日交易，在正常的交易时间有 194 个价格变化. 图 5-15 显示了 $\ln(\Delta t_i)$，N_i，D_i 和 S_i 的直方图. D_i 的数据在“上升”与“下降”运动中大约是平等分布的. 只有极少数交易导致了大于 1 个变动价位的价格变化；事实上，有七个具有两个变动价位的变化，有一个具有三个变动价位的变化. 利用 MCMC 方法(详见第 10 章)，我们对数据得到下面的模型. 描述的估计及其标准差是 9 500 次迭代中 MCMC 抽取的后验均值和标准差. 价格变化之间的时间持续期模型为

$$\ln(\Delta t_i) = 4.023 + 0.032\ln(\Delta t_{i-1}) - 0.025 S_{i-1} + 1.403\epsilon_i$$

210

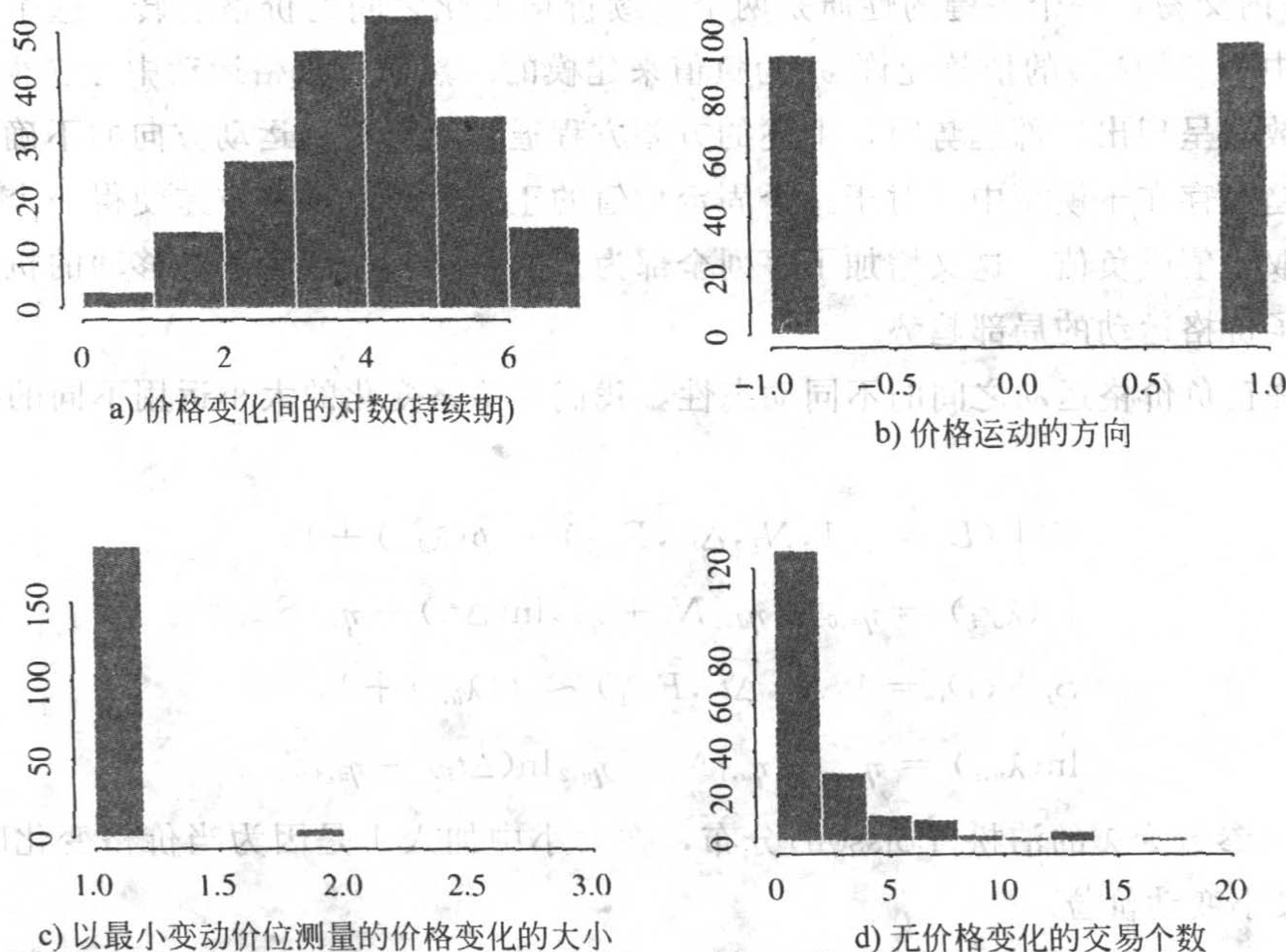
其中系数的标准差分别为 0.415，0.073，0.384 和 0.073. 拟合的模型显示在时间持续期中没有动态相依性. 对变量 N_i，我们有

$$Pr(N_i > 0 \mid \Delta t_i, F_{i-1}) = \text{logit}[-0.637 + 1.740\ln(\Delta t_i)],$$

其中估计的标准差分别为 0.238 和 0.248. 这样，在时间间隔(t_{i-1}, t_i)上无价格变化的交易数量正向依赖于间隔的长度. 当 $N_i>0$ 时，其大小为

$$N_i \mid (N_i > 0, \Delta t_i, F_{i-1}) \sim 1 + g(\lambda_i), \lambda_i = \frac{\exp[0.178 - 0.910\ln(\Delta t_i)]}{1 + \exp[0.178 - 0.910\ln(\Delta t_i)]}$$

其中估计的标准差分别为 0.246 和 0.138. $\ln(\Delta t_i)$的显著的负系数意味着 N_i 与持续期 Δt_i 的长度是正相关的，因为大的 $\ln(\Delta t_i)$预示着一个小的 λ_i，从而隐含了 N_i 以更高的概率取大的值，可以参见方程(5.27)中的几何分布.



211

图 5-15 IBM 股票在 1990 年 11 月 21 日的日内交易数据的直方图

对 D_i 拟合的模型为

$$\mu_i = 0.049 - 0.840 D_{i-1} - 0.004 \ln(\Delta t_i),$$
$$\ln(\sigma_i) = 0.244 \, | D_{i-1} + D_{i-2} + D_{i-3} + D_{i-4} |$$

其中均值方程中参数的标准差分别为 0.129，0.132 和 0.082，而方差方程中参数的标准差为 0.182. 价格逆转可以从 D_{i-1} 的高度显著的负系数中清楚地显示出来. 方差方程的边际显著参数也正好是所期望的. 最后，对价格变化大小的拟合模型为

$$\ln(\lambda_{d,i}) = 1.024 - 0.327 N_i + 0.412 \ln(\Delta t_i) - 4.474 S_{i-1},$$
$$\ln(\lambda_{u,i}) = -3.683 - 1.542 N_i + 0.419 \ln(\Delta t_i) + 0.921 S_{i-1}$$

其中“下降大小”参数的标准差分别为 3.350，0.319，0.599 和 3.188，而“上升大小”的参数的标准差分别为 1.734，0.976，0.453 和 1.459. 上述两个方程的令人感兴趣的估计是 N_i 的系数的估计值为负数. 大 N_i 意味着在时间间隔 (t_{i-1}, t_i) 中有更多的交易没有价格变化，这可以看作是时间间隔 (t_{i-1}, t_i) 中没有可以利用的新信息的证据. 因此，t_i 时刻价格变化的大小应该比较小. 泊松分布中的小 $\lambda_{u,i}$ 与 $\lambda_{d,i}$ 精确地说明了这一点.

总之，假定某天的 194 个观测值样本可能不包含关于 IBM 股票交易动态的充分信息，但是拟合的模型看上去提供了一些合理的结果. McCulloch 和 Tsay(2000) 把 PCD 模型扩展到一个分级结构框架，来处理从 1990 年 11 月 1 日至 1991 年 1 月 31 日之间的所有 63 个交易日的数据. 有超过 19 000 个观测值，许多参数估计在这个扩展模型中都是显著的. 例如，在时间持

续期模型中$\ln(\Delta t_{i-1})$的系数的全部估计从0.04到0.1变化，尽管比较小但是很显著.

最后，利用交易数据来检验微观结构理论要求仔细地指定所用变量. 还要求对市场运作的方式与数据收集的更深刻的理解. 然而，本章中讨论的经济计量模型的思想是有用的，在高频数据分析中广泛地得到应用.

附录 A　一些概率分布的回顾

指数分布

称随机变量 X 有参数为$\beta(\beta>0)$的指数分布，如果其概率密度函数(pdf)由下式给出 [212]

$$f(x \mid \beta)=\begin{cases}\dfrac{1}{\beta}\mathrm{e}^{-x/\beta} & 若\ x \geqslant 0 \\ 0 & 其他\end{cases}$$

这样一个分布表示为 $X \sim \exp(\beta)$. 我们有 $E(X)=\beta$，$\operatorname{Var}(X)=\beta^2$. X 的累积分布函数(CDF)为

$$f(x \mid \beta)=\begin{cases}0 & 若\ x<0 \\ 1-\mathrm{e}^{-x/\beta} & 若\ x \geqslant 0\end{cases}$$

当$\beta=1$时，称 X 有标准指数分布.

伽马函数

对$\kappa>0$，伽马函数$\Gamma(\kappa)$定义为

$$\Gamma(\kappa)=\int_0^{\infty} x^{\kappa-1}\mathrm{e}^{-x}\mathrm{d}x$$

伽马函数的重要性质为：

1)对任意的$\kappa>1$，$\Gamma(\kappa)=(\kappa-1)\Gamma(\kappa-1)$.

2)对任何正整数值m，$\Gamma(m)=(m-1)!$.

3)$\Gamma\left(\dfrac{1}{2}\right)=\sqrt{\pi}$.

积分

$$\Gamma(y \mid \kappa)=\int_0^{y} x^{\kappa-1}\mathrm{e}^{-x}\mathrm{d}x, y>0$$

是一个不完全的伽马函数，它的值在文献中已用表列出. 估计不完全伽马函数的计算机程序现在是可以得到的.

伽马分布

称随机变量 X 有参数为κ和$\beta(\kappa>0, \beta>0)$的伽马分布，如果其 pdf 由下式给出

$$f(x \mid \kappa, \beta)=\begin{cases}\dfrac{1}{\beta^{\kappa}\Gamma(\kappa)}x^{\kappa-1}\mathrm{e}^{-x/\beta} & 若\ x \geqslant 0 \\ 0 & 其他\end{cases}$$

通过变量代换$y=x/\beta$，很容易可以得到 X 的各阶矩：

[213]

$$E(X^m)=\int_0^\infty x^m f(x\mid\kappa,\beta)\mathrm{d}x=\frac{1}{\beta^\kappa\Gamma(\kappa)}\int_0^\infty x^{\kappa+m-1}\mathrm{e}^{-x/\beta}\mathrm{d}x$$
$$=\frac{\beta^m}{\Gamma(\kappa)}\int_0^\infty y^{\kappa+m-1}\mathrm{e}^{-y}\mathrm{d}y=\frac{\beta^m\Gamma(\kappa+m)}{\Gamma(\kappa)}$$

特殊地，X 的均值、方差分别为 $E(X)=\kappa\beta$，$\mathrm{Var}(X)=\kappa\beta^2$. 当 $\beta=1$ 时，该分布称为参数 κ 的标准伽马分布. 利用记号 $G\sim\mathrm{Gamma}(\kappa)$表示 G 服从一个参数为 κ 的标准伽马分布. G 的各阶矩为

$$E(G^m)=\frac{\Gamma(\kappa+m)}{\Gamma(\kappa)}\quad m>0 \tag{5.54}$$

韦布尔分布

称一个随机变量 X 有参数为 α、$\beta(\alpha>0,\ \beta>0)$的韦布尔分布，如果其 pdf 为

$$f(x\mid\alpha,\beta)=\begin{cases}\dfrac{\alpha}{\beta^\alpha}x^{\alpha-1}\mathrm{e}^{-(x/\beta)^\alpha} & 若\ x\geqslant 0\\ 0 & 若\ x<0\end{cases}$$

这里 β 和 α 分别为分布的尺度参数和形状参数. X 的均值、方差分别为

$$E(X)=\beta\Gamma\left(1+\frac{1}{\alpha}\right),\mathrm{Var}(X)=\beta^2\left\{\Gamma\left(1+\frac{2}{\alpha}\right)-\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2\right\}$$

X 的 CDF 为

$$F(x\mid\alpha,\beta)=\begin{cases}0 & 若\ x<0\\ 1-\mathrm{e}^{-(x/\beta)^\alpha} & 若\ x\geqslant 0\end{cases}$$

当 $\alpha=1$ 时，韦布尔分布简化为指数分布.

定义 $Y=X/\left[\beta\Gamma\left(1+\frac{1}{\alpha}\right)\right]$. 我们有 $E(Y)=1$，而且 Y 的 pdf 为

$$f(y\mid\alpha)=\begin{cases}\alpha\left[\Gamma\left(1+\dfrac{1}{\alpha}\right)\right]^\alpha y^{\alpha-1}\exp\left\{-\left[\Gamma\left(1+\dfrac{1}{\alpha}\right)y\right]^\alpha\right\} & 若\ y\geqslant 0\\ 0 & 其他\end{cases} \tag{5.55}$$

这里，由于标准化，所以尺度参数消失了. 标准化的韦布尔分布的 CDF 为

$$F(y\mid\alpha)=\begin{cases}0 & 若\ y<0\\ 1-\exp\left\{-\left[\Gamma\left(1+\dfrac{1}{\alpha}\right)y\right]^\alpha\right\} & 若\ y>0,\end{cases}$$

并且我们有 $E(Y)=1$，$\mathrm{Var}(Y)=\Gamma\left(1+\frac{2}{\alpha}\right)\Big/\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2-1$. 对于带韦布尔新息的持续期模

[214] 型，最大似然估计中利用的就是上述的 pdf.

广义伽马分布

称随机变量 X 有参数为 α，β，$\kappa(\alpha>0,\ \beta>0,\ \kappa>0)$的广义伽马分布，如果它的 pdf 由下式给出

$$f(x \mid \alpha,\beta,\kappa) = \begin{cases} \dfrac{\alpha x^{\kappa\alpha-1}}{\beta^{\kappa\alpha}\Gamma(\kappa)}\exp\left[-\left(\dfrac{x}{\beta}\right)^{\alpha}\right] & 若 x \geqslant 0 \\ 0 & 其他 \end{cases}$$

其中 β 是尺度参数，α、κ 为形状参数．这个分布可以写为

$$G = \left(\frac{X}{\beta}\right)^{\alpha}$$

这里 G 是参数为 κ 的标准伽马随机变量．X 的 pdf 可以通过变量代换的方法由 G 得到．类似地，X 的矩可以从方程(5.54)中 G 的矩得到：

$$E(X^m) = E[(\beta G^{1/\alpha})^m] = \beta^m E(G^{m/\alpha}) = \beta^m \frac{\Gamma\left(\kappa + \dfrac{m}{\alpha}\right)}{\Gamma(\kappa)} = \frac{\beta^m \Gamma\left(\kappa + \dfrac{m}{\alpha}\right)}{\Gamma(\kappa)}$$

当 $\kappa=1$ 时，广义伽马分布简化为韦布尔分布．这样，指数分布和韦布尔分布都是广义伽马分布的特殊情况．

广义伽马分布的期望为 $E(X)=\beta\Gamma\left(\kappa+\dfrac{1}{\alpha}\right)/\Gamma(\kappa)$．在持续期模型中，我们需要一个带单位期望的分布．因此，定义随机变量 $Y=\lambda X/\beta$，这里 $\lambda=\Gamma(\kappa)/\Gamma\left(\kappa+\dfrac{1}{\alpha}\right)$，则有 $E(Y)=1$，Y 的 pdf 为

$$f(y \mid \alpha,\kappa) = \begin{cases} \dfrac{\alpha y^{\kappa\alpha-1}}{\lambda^{\kappa\alpha}\Gamma(\kappa)}\exp\left[-\left(\dfrac{y}{\lambda}\right)^{\alpha}\right] & 若 y > 0 \\ 0 & 其他 \end{cases} \tag{5.56}$$

这里 β 仍然是尺度参数，$\lambda=\Gamma(\kappa)/\Gamma\left(\kappa+\dfrac{1}{\alpha}\right)$.

附录 B　危险率函数

对持续期建模时一个有用的概念是由分布函数隐含的危险率函数．对随机变量 X，生存函数定义为

$$S(x) \equiv P(X > x) = 1 - P(X \leqslant x) = 1 - \text{CDF}(x), \qquad x > 0$$

这给出了服从 X 的分布的每个事物 x 在时刻 t 生存的概率．X 的危险率函数(或强度函数)定义为 [215]

$$h(x) = \frac{f(x)}{S(x)} \tag{5.57}$$

其中 $f(.)$和 $S(.)$分别是 X 的 pdf 和生存函数．

例 5.6　对参数为 α、β 的韦布尔分布，生存函数与危险率函数分别为

$$S(x \mid \alpha,\beta) = \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], \quad h(x \mid \alpha,\beta) = \frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}, \qquad x > 0$$

特别地，当 $\alpha=1$ 时，我们有 $h(x \mid \beta)=1/\beta$. 因此，对指数分布而言，其危险率函数是常数，对韦布尔分布，危险率函数是单调的. 如果 $\alpha>1$，那么危险率函数是单调递增的，如果 $\alpha<1$，那么危险率函数是单调递减的. 对广义伽马分布，生存函数以及危险率函数都涉及不完全伽马函数. 然而危险率函数可以有各种不同的模式，包括 U 型以及倒 U 型. 这样，广义伽马分布提供了对股票交易的持续期建模的一个灵活方法.

对标准化的韦布尔分布，生存函数以及危险率函数为

$$S(y \mid \alpha)=\exp\left\{-\left[\Gamma\left(1+\frac{1}{\alpha}\right)y\right]^{\alpha}\right\},$$

$$h(y \mid \alpha)=\alpha\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^{\alpha}y^{\alpha-1},\qquad y>0.$$

附录 C 持续期模型的一些 RATS 程序

运用的数据是 IBM 股票从 1990 年 11 月 1 日至 11 月 9 日每日交易的调整后的时间持续期. 文件名为"ibmlto5. dat"，共有 3 534 个观测.

A. 估计 WACD(1，1)模型的程序

216

```
all 0 3534:1
open data ibm1to5.dat
data(org=obs) / x r1
set psi = 1.0
nonlin a0 a1 b1 al
frml gvar = a0+a1*x(t-1)+b1*psi(t-1)
frml gma = %LNGAMMA(1.0+1.0/al)
frml gln =al*gma(t)+log(al)-log(x(t)) $
 +al*log(x(t)/(psi(t)=gvar(t)))-(exp(gma(t))*x(t)/psi(t))**al
smpl 2 3534
compute a0 = 0.2, a1 = 0.1, b1 = 0.1, al = 0.8
maximize(method=bhhh,recursive,iterations=150) gln
set fv = gvar(t)
set resid = x(t)/fv(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

B. 估计 GACD(1，1)模型的程序

```
all 0 3534:1
open data ibm1to5.dat
data(org=obs) / x r1
set psi = 1.0
nonlin a0 a1 b1 al ka
frml cv = a0+a1*x(t-1)+b1*psi(t-1)
frml gma = %LNGAMMA(ka)
frml lam = exp(gma(t))/exp(%LNGAMMA(ka+(1.0/al)))
frml xlam = x(t)/(lam(t)*(psi(t)=cv(t)))
```

```
frml gln =-gma(t)+log(al/x(t))+ka*al*log(xlam(t))-(xlam(t))**al
smpl 2 3534
compute a0 - 0.238, a1 = 0.075, b1 = 0.857, al = 0.5, ka = 4.0
nlpar(criterion=value,cvcrit=0.00001)
maximize(method=bhhh,recursive,iterations=150) gln
set fv = cv(t)
set resid = x(t)/fv(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

C. 估计 Tar-WACD(1，1)模型的程序，指定的门限为 3.79

```
all 0 3534:1
open data ibm1to5.dat
data(org=obs) / x rt
set psi = 1.0
nonlin a1 a2 al b0 b2 bl
frml u = ((x(t-1)-3.79)/abs(x(t-1)-3.79)+1.0)/2.0
frml cp1 = a1*x(t-1)+a2*psi(t-1)
frml gma1 = %LNGAMMA(1.0+1.0/al)
frml cp2 = b0+b2*psi(t-1)
frml gma2 = %LNGAMMA(1.0+1.0/bl)
frml cp = cp1(t)*(1-u(t))+cp2(t)*u(t)
frml gln1 =al*gma1(t)+log(al)-log(x(t)) $
 +al*log(x(t)/(psi(t)=cp(t)))-(exp(gma1(t))*x(t)/psi(t))**al
frml gln2 =bl*gma2(t)+log(bl)-log(x(t)) $
 +bl*log(x(t)/(psi(t)=cp(t)))-(exp(gma2(t))*x(t)/psi(t))**bl
frml gln = gln1(t)*(1-u(t))+gln2(t)*u(t)
smpl 2 3534
compute a1 = 0.2, a2 = 0.85, al = 0.9
compute b0 = 1.8, b2 = 0.5, bl = 0.8
maximize(method=bhhh,recursive,iterations=150) gln
set fv = cp(t)
set resid = x(t)/fv(t)
set residsq = resid(t)*resid(t)
cor(qstats,number=20,span=10) resid
cor(qstats,number=20,span=10) residsq
```

[217]

练习题

1. 令 r_t 为资产在 t 时刻的对数收益率．假定$\{r_t\}$是均值为 0.05、方差为 1.5 的高斯白噪声序列．还假定在每个时间点交易发生的概率为 40%，并且交易是否发生与 r_t 独立．用 r_t^o 表示观测到的收益率．问：r_t^o 是前后相关吗？如果回答是肯定的，请计算 r_t^o 的前 3 步延迟自相关系数．

2. 令 P_t 是观测到的资产的市场价格，它通过方程(5.9)与资产的基本价值 P_t^* 相联系．假定 $\Delta P_t^* = P_t^* - P_{t-1}^*$ 是一个均值为 0、方差为 1.0 的高斯白噪声序列，并假定买卖报价差为 2 个最小变动价位．问：当最小变动价位是 1/8 美元时，价格变化序列 $\Delta P_t = P_t - P_{t-1}$ 的 1 步

延迟自相关系数是多少？当最小变动价位是 1/16 美元时，价格变化的 1 步延迟自相关系数是多少？

3. 文件“ibm-d2-dur. dat”包含了 IBM 股票 1990 年 11 月 2 日交易之间调整的持续期. 这个文件有 3 列，含有天、从午夜开始以秒测量的交易时刻以及调整持续期.
   (a)对调整持续期建立一个 EACD 模型，并检验拟合的模型.
   (b)对调整持续期建立一个 WACD 模型，并检验拟合的模型.
   (c)对调整持续期建立一个 GACD 模型，并检验拟合的模型.
   (d)比较前面的三个持续期模型.
4. 文件“mmm9912-dtp. dat”包含了 3M 公司的股票在 1999 年 12 月的交易数据. 有 3 列：这个月的天、从午夜开始以秒测量的交易时刻、交易价格. 排除了东部时间 4：00 以后发生的交易.
   (a)3M 股票交易具有日模式吗？你可以构造一个时间序列 n_t，用它表示每 5 分钟时间间隔内的交易数量来回答这个问题.
   (b)利用价格序列来确认 3M 股票的日内交易中买卖报价弹性的存在.
   (c)以最小变动价位 1/16 美元的倍数将价格变化的频率列表. 你可以将变化为最小变动价位的 5 倍或超过 5 倍的分成一类，变化为最小变动价位的 −5 倍或超出 −5 倍的分成另一类.

[218]

5. 再次考虑 3M 股票在 1999 年 12 月的交易数据.
   (a)利用此数据构造日内每 5 分钟的对数收益率序列，利用每 5 分钟间隔内所有交易价格的简单平均作为这个间隔的股票价格. 问：此序列是前后相关的吗？可以利用样本自相关函数的前 10 步延迟值的 Ljung-Box 统计量来检验这个假设.
   (b)在正常的交易日共有 77 个每 5 分钟收益率. 某些研究者建议可以用日内的每 5 分钟收益率的平方和作为日波动率的一个度量. 采用这个方法，并计算 3M 股票在 1999 年 12 月的对数收益率的日波动率. 讨论用这样的方法来估计日波动率的有效性.
6. 文件“mmm9912-adur. dat”包含了 3M 股票在 1999 年 12 月调整的日交易持续期. 在一个交易日内有 39 个每 10 分钟的时间间隔. 令 d_i 为 1999 年 12 月所有的交易日内第 i 个 10 分钟间隔的所有对数持续期的平均. 定义调整的持续期为 $t_j/\exp(d_i)$，其中 j 是在第 i 个 10 分钟间隔内. 注意到更多复杂的方法可以用来调整交易持续期的日内模式. 这里，我们仅仅使用局部平均.
   (a)在调整的持续期序列中有日模式吗？为什么？
   (b)利用指数新息对调整的序列构造一个持续期模型，并检验拟合的模型.
   (c)利用韦布尔新息对调整的序列构造一个持续期模型，并检验拟合的模型.
   (d)利用广义伽马新息对调整的序列构造一个持续期模型，并检验拟合的模型.
   (e)比较并评论前面构造的 3 个持续期模型.

参考文献

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press: New Jersey.

Cho, D., Russell, J. R., Tiao, G. C., and Tsay, R. S. (2000), "The magnet effect of price limits: Evidence from high frequency data on Taiwan stock exchange," Working paper, Graduate School of Business, University of Chicago.

Engle, R. F., and Russell, J. R. (1998), "Autoregressive conditional duration: A new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Ghysels, E. (2000), "Some econometric recipes for high-frequency data cooking," *Journal of Business and Economic Statistics*, 18, 154–163.

Hasbrouck, J. (1992), *Using the TORQ database*, Stern School of Business, New York University.

Hasbrouck, J. (1999), "The dynamics of discrete bid and ask quotes," *Journal of Finance*, 54, 2109–2142.

Hauseman, J., Lo, A., and MacKinlay, C. (1992), "An ordered probit analysis of transaction stock prices," *Journal of Financial Economics*, 31, 319–379. 219

Lo, A., and MacKinlay, A. C. (1990), "An econometric analysis of nonsynchronous trading," *Journal of Econometrics*, 45, 181–212.

McCulloch, R. E., and Tsay, R. S. (2000), "Nonlinearity in high frequency data and hierarchical models," Working paper, Graduate School of Business, University of Chicago.

Roll, R. (1984), "A simple implicit measure of the effective bid-ask spread in an efficient market," *Journal of Finance*, 39, 1127–1140.

Rydberg, T. H., and Shephard, N. (1998), "Dynamics of trade-by-trade price movements: decomposition and models," Working paper, Nuffield College, Oxford University.

Stoll, H., and Whaley, R. (1990), "Stock market structure and volatility," *Review of Financial Studies*, 3, 37–71.

Wood, R. A. (2000), "Market microstructure research databases: History and projections," *Journal of Business & Economic Statistics*, 18, 140–145.

Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001), "A nonlinear autoregressive conditional duration model with applications to financial transaction data," *Journal of Econometrics* (to appear).

Zhang, M. Y., Russell, J. R., and Tsay, R. S. (2001b), "Determinants of bid and ask quotes and implications for the cost of trading," Working paper, Graduate School of Business, University of Chicago. 220

第 6 章　连续时间模型及其应用

金融资产的价格随时间变化，形成一个随机过程．随机过程是用来描述一个随机变量随时间变化的统计术语．观测到的价格是随机过程的一个实现．随机过程的理论是对观测到的价格进行分析和作出统计推断的基础．

有两种随机过程模型可用来对资产价格建模．第一种称为离散时间随机过程，其中价格变化发生在离散的时间点．前面章节讨论的过程都属于这个类型．例如，纽约证券交易所 IBM 股票的日收盘价组成一个离散时间的随机过程，这里价格只在一个交易日的收盘时刻发生变化．一个交易日内的价格运动不一定与观测到的日价格有关．第二种类型是连续时间的随机过程，尽管价格只在离散时间点上可以观测到，但价格变化是连续的．可以将价格认为是股票的“真实价值”，总是存在而且是时变的．

对两种类型的随机过程，价格可以是连续的或者离散的．连续的价格可以假定为任何正实数，而离散价格只能假定可能取值构成一个可数集合．假定资产价格是一个连续时间的随机过程．如果价格是一个连续的随机变量，那么我们就有一个连续时间的连续过程．如果价格本身是离散的，那么我们有一个连续时间的离散过程．对于离散时间过程也可以有类似的分类．第 5 章中的价格变化序列是离散时间的离散过程的一个例子．

在本章中，我们将资产价格看作是一个连续时间的连续随机过程．我们的目的是引进对金融资产和价格运动建模需要的统计理论和工具．在本章的开始，我们介绍本章中使用的一些股票期权的技术．在 6.2 节，我们对布朗运动(Brownian motion)(也称为维纳过程(Wiener process))作一个简要的介绍．然后我们讨论一些扩散方程以及随机方程，包括著名的伊藤引理(Ito's lemma)．大多数期权定价公式是在资产价格服从一个扩散方程的假设下推导出来的．我们利用 Black-Scholes(以下简称 B-S)公式来表示这种推导．最后，为了处理稀有事件(比如一个利润警告)引起的价格变化，我们也研究一些简单的带跳跃的扩散模型．

221

如果资产价格服从一个扩散方程，那么附随于资产的期权价格可以利用对冲(hedging)方法推导出来．然而，有跳跃的时候，市场变得不完全从而不存在完全的期权对冲．此时期权的价格或者利用跳跃风险的分散性来估价，或者定义一个风险的度量，然后选择一个价格和对冲来最小化这个风险．对于随机过程在衍生产品定价过程中的基本应用，参见 Cox 和 Rubinstein (1985) 以及 Hull(1997)．

6.1　期权

股票期权是一个金融契约，它赋予持有者在特定的日期以一个指定的价格交易一个指定的普通股票的特定份额的权利，它有两种类型．看涨期权赋予持有者买某种标的资产的权利，其

正式的定义见第 3 章. 看跌期权赋予持有者卖某种标的资产的权利. 契约中指定的价格称为敲定价格或执行价格. 契约规定的日期称为到期日或截止日. 美式期权可以在到期日之前的任何时刻执行，欧式期权只能在到期日执行.

股票期权的价值依赖于标的资产的价值. 令 K 表示敲定价格，P 是股价. 买权当 $P>K$ 时，赚钱；当 $P=K$ 时，不赔不赚；当 $P<K$ 时，赔钱. 卖权当$P<K$时，赚钱；当 $P=K$ 时，不赔不赚；当 $P>K$ 时，赔钱. 一般而言，一个期权如果立即执行，将对持有者导致正的现金流时是赚钱的. 期权如果立即执行，将对股票持有者导致负现金流时是赔钱的. 最后，期权如果立即执行，将导致 0 现金流时是不赔不赚的. 显而易见，只有赚钱的期权在实际中会被执行. 要得到更多关于期权的描述，参见 Hull(1997).

6.2 一些连续时间的随机过程

在数理统计中，连续时间的连续随机过程是定义在一个概率空间(Ω, F, P)上，其中 Ω 是非空空间，F 是一个包含 Ω 的所有子集的 σ 域，P 是概率测度；详见 Billingsley(1986)第 1 章. 过程可以写为$\{x(\eta, t)\}$，其中 t 表示时间，在$[0, \infty)$上连续. 对于给定的 t，$x(\eta, t)$是一个实值连续随机变量(即从 Ω 到实直线上的一个映射)，η 是 Ω 的元素. 对于资产在时刻 t 的价格，$x(\eta, t)$的取值范围是非负实数的集合. 对于一个给定的 η，$\{x(\eta, t)\}$是一个时间序列，其值依赖于时刻 t. 为了简便，我们将连续时间随机过程写为$\{x_t\}$，可以理解为，对给定的 t，x_t 是一个随机变量. 在文献中，一些作者为了强调 t 是连续的，运用 $x(t)$来代替 x_t. 然而我们利用相同的记号 x_t，但是称为一个连续时间随机过程. 222

6.2.1 维纳过程

在离散时间经济计量模型中，我们假定抖动形成一个白噪声过程它是不可预测的. 连续时间模型中抖动的对应部分是什么？答案是一个维纳过程(也称为标准布朗运动)的增量. 有很多方式定义一个维纳过程$\{w_t\}$，我们利用与在时间上的小增量 Δt 相关的小变化 $\Delta w_t=w_{t+\Delta t}-w_t$ 的一个简单方法. 一个连续时间随机过程$\{w_t\}$是一个维纳过程，如果它满足

1) $\Delta w_t=\epsilon\sqrt{\Delta t}$，其中 ϵ 是一个标准正态随机变量，

2) Δw_t 与 w_j 独立，对于所有的 $j\leqslant t$.

第二个条件是一个马尔可夫性，说明在当前值 w_t 的条件下，过程过去的任何信息 $w_j(j<t)$ 与将来值 $w_{t+l}(l>0)$是不相关的. 由这个性质很容易看出，对任意两个不相交的时间段 Δ_1 与 Δ_2，增量 $w_{t_1+\Delta_1}-w_{t_1}$ 与增量 $w_{t_2+\Delta_2}-w_{t_2}$ 是独立的. 在金融中，这个马尔可夫性与弱式有效市场是相关的.

由第一个条件，Δw_t 是均值为 0、方差为 Δt 的正态分布. 也就是，$\Delta w_t\sim N(0, \Delta t)$，这里～表示概率分布. 下面考虑过程 w_t. 我们假定过程开始于 $t=0$，初始值为 w_0，这个值是固定的，而且通常设定为 0. 从而 w_t-w_0 可以看作是许多微小增量的和. 更具体地，定义 $T=\frac{t}{\Delta t}$，这里 Δt 为一个很小的正增量，则

$$w_t - w_0 = w_{T\Delta t} - w_0 = \sum_{i=1}^{T} \Delta w_i = \sum_{i=1}^{T} \varepsilon_i \sqrt{\Delta t},$$

其中 $\Delta w_i = w_{i\Delta t} - w_{(i-1)\Delta t}$. 因为诸 ε_i 是独立的，所以我们有

$$E(w_t - w_0) = 0, \quad \mathrm{Var}(w_t - w_0) = \sum_{i=1}^{T} \Delta t = T\Delta t = t$$

这样，w_t 从 0 时刻到 t 时刻的增量是均值为 0 方差为 t 的正态分布. 规范地表示为，对于维纳过程 w_t，我们有 $w_t - w_0 \sim N(0, t)$. 这说明一个维纳过程的方差是随时间间隔的长度线性增加的.

223

图 6-1 显示的是单位时间区间[0, 1]上的四个模拟的维纳过程. 它们是由在统计学文献中的 Donsker 定理的一个简单形式（参见 Donsker（1951）或 Bilingsley（1968））得到的，取 $n=3000$.

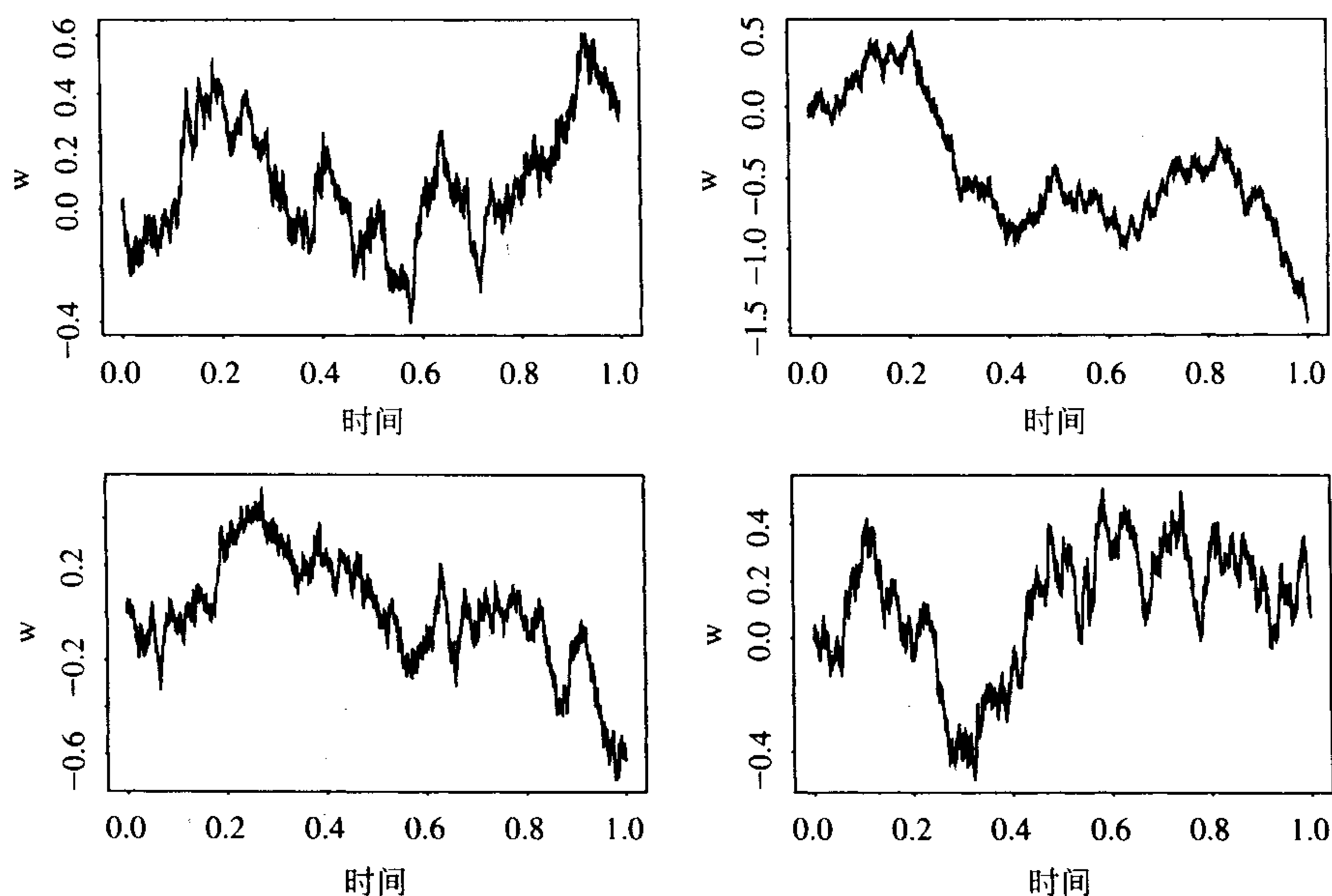


图 6-1 四个模拟的维纳过程

Donsker 定理

假定 $\{z_i\}_{i=1}^n$ 是一个相互独立的标准正态随机变量的序列. 对于任意 $t \in [0,1]$，令 $[nt]$ 表示 nt 的整数部分. 定义 $w_{n,t} = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} z_i$. 则当 n 趋于无穷时，$w_{n,t}$ 依分布收敛到一个在[0,1]区间上的维纳过程 w_t.

这四幅图都开始于 $w_0 = 0$，但是当时间递增时漂移开来，说明维纳过程的方差是随时间递增的. 可以用一个从[0, 1)到[0, ∞)的一个简单的时间变换来得到 $t \in [0, \infty)$ 上模拟的维纳过程.

注释： 概率空间(Ω, F, P)上的布朗运动 w_t 正式定义为：对于 $t \geqslant 0$，是一个具有独

立平稳增量的实值连续随机过程．换句话说，w_t 满足

1)连续性：从 t 到 w_t 的映射关于概率测度 P 是几乎处处连续的；

2)独立增量：如果 $s \leqslant t$，则对于所有的 $v \leqslant s$，$w_t - w_s$ 与 w_v 独立； 224

3)平稳增量：如果 $s \leqslant t$，则 $w_t - w_s$ 与 $w_{t-s} - w_0$ 具有同样的概率分布．

可以证明增量 $w_t - w_s$ 的概率分布是均值为 $\mu(t-s)$，方差为 $\sigma^2(t-s)$ 的正态分布．而且，对于任意给定的时间指标 $0 \leqslant t_1 < t_2 < \cdots < t_k$，随机向量 $(w_{t_1}, w_{t_2}, \cdots, w_{t_k})$ 服从一个多元正态分布．最后，如果 $w_0 = 0$ 几乎处处成立，$\mu = 0$ 且 $\sigma^2 = 1$，则布朗运动是标准的．

注释：布朗运动的一个重要性质是它们的路径是几乎处处不可微的．换句话说，对一个标准布朗运动 w_t，对于 Ω 中除了满足 $P(\Omega_1) = 0$ 的子集合 $\Omega_1 \subset \Omega$ 以外的所有元素，$\mathrm{d}w_t/\mathrm{d}t$ 都是不存在的．因此，当考虑资产的时间价值时，我们不能运用通常的积分求和来处理涉及到标准布朗运动的积分，必须寻求另外一种方法．这就是下一节中讨论伊藤积分的目的．

6.2.2 一般的维纳过程

维纳过程是一个特殊的随机过程，具有 0 漂移率以及与时间间隔的长度成比例的方差．这意味着期望的变化率为 0，而且方差的变化率为 1．实际中，随机过程的均值和方差可以以一种更加复杂的方式随时间演变，因此需要随机过程的进一步一般化．为了这个目的，我们考虑期望漂移率为 μ，方差变化率为 σ^2 的一般维纳过程．将这个过程用 x_t 来表示，并且用记号 $\mathrm{d}y$ 表示变量 y 的一个微小变化，则 x_t 的模型为

$$\mathrm{d}x_t = \mu \mathrm{d}t + \sigma \mathrm{d}w_t \tag{6.1}$$

其中 w_t 是一个维纳过程．如果我们考虑方程(6.1)的离散形式，那么对于从 0 到 t 的增量

$$x_t - x_0 = \mu t + \sigma \epsilon \sqrt{t}$$

因此，

$$E(x_t - x_0) = \mu t, \quad \mathrm{Var}(x_t - x_0) = \sigma^2 t,$$

结果说明 x_t 的增量对期望增长率为 μ，对方差的增长率为 σ^2．在文献中，方程(6.1)中的 μ 与 σ 称为一般维纳过程 x_t 的漂移参数和波动率参数．

6.2.3 伊藤过程

一般维纳过程的漂移参数和波动率参数都是不随时间变化的．如果进一步扩展模型，允许 μ 和 σ 是随机过程 x_t 的函数，那么我们就有一个伊藤过程．具体地，过程 x_t 是一个伊藤过程，如果它满足 225

$$\mathrm{d}x_t = \mu(x_t, t)\mathrm{d}t + \sigma(x_t, t)\mathrm{d}w_t \tag{6.2}$$

其中 w_t 是一个维纳过程．这个过程在数理金融中起着很重要的作用，可以写为

$$x_t = x_0 + \int_0^t \mu(x_s, s)\mathrm{d}s + \int_0^t \sigma(x_s, s)\mathrm{d}w_s$$

其中 x_0 表示过程在 0 时刻的初始值，右面的最后一项是一个随机积分．方程(6.2)称为一个随

机扩散方程，$\mu(x_t, t)$和$\sigma(x_t, t)$分别是漂移函数和扩散函数.

维纳过程是一个特殊的伊藤过程，因为它满足方程(6.2)中取$\mu(x_t, t)=0$且$\sigma(x_t, t)=1$的情形.

6.3 伊藤引理

金融中利用连续时间模型时，通常假定资产的价格是一个伊藤过程. 因此，为了导出金融衍生资产的价格，需要利用伊藤积分. 在本节中，我们主要回顾一下伊藤引理，将它看作是微积分中微分运算的一个自然推广. 伊藤引理是随机积分的基础.

6.3.1 微分回顾

令$G(x)$表示x的可微函数. 利用泰勒展开，我们有

$$\Delta G \equiv G(x+\Delta x) - G(x) = \frac{\partial G}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}(\Delta x)^2 + \frac{1}{6}\frac{\partial^3 G}{\partial x^3}(\Delta x)^3 + \cdots$$

当$\Delta x \to 0$时取极限，并且忽略Δx的高阶项，我们有

226

$$\mathrm{d}G = \frac{\partial G}{\partial x}\mathrm{d}x$$

当G为x和y的函数时，我们有

$$\Delta G = \frac{\partial G}{\partial x}\Delta x + \frac{\partial G}{\partial y}\Delta y + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}(\Delta x)^2 + \frac{\partial^2 G}{\partial x \partial y}\Delta x \Delta y + \frac{1}{2}\frac{\partial^2 G}{\partial y^2}(\Delta y)^2 + \cdots$$

当$\Delta x \to 0$且$\Delta y \to 0$时取极限，我们有

$$\mathrm{d}G = \frac{\partial G}{\partial x}\mathrm{d}x + \frac{\partial G}{\partial y}\mathrm{d}y.$$

6.3.2 随机微分

下面转到考虑G为x_t和t的可微函数的情况，其中x_t是一个伊藤过程. 泰勒展开变为

$$\Delta G = \frac{\partial G}{\partial x}\Delta x + \frac{\partial G}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}(\Delta x)^2 + \frac{\partial^2 G}{\partial x \partial t}\Delta x \Delta t + \frac{1}{2}\frac{\partial^2 G}{\partial t^2}(\Delta t)^2 + \cdots \tag{6.3}$$

伊藤过程的离散形式为

$$\Delta x = \mu \Delta t + \sigma \varepsilon \sqrt{\Delta t} \tag{6.4}$$

这里为了简便，我们省略了μ和σ的变元，并记$\Delta x = x_{t+\Delta t} - x_t$. 由方程(6.4)，我们有

$$(\Delta x)^2 = \mu^2(\Delta t)^2 + \sigma^2\varepsilon^2\Delta t + 2\mu\sigma\varepsilon(\Delta t)^{3/2} = \sigma^2\varepsilon^2\Delta t + H(\Delta t) \tag{6.5}$$

其中$H(\Delta t)$表示Δt的高阶项. 这个结果说明$(\Delta x)^2$包含了与Δt同阶的项，当我们对$\Delta t \to 0$取极限时，此项不能忽略. 然而，方程(6.5)右面的第一项有一些好的性质：

$$E(\sigma^2\varepsilon^2\Delta t) = \sigma^2\Delta t$$

$$\mathrm{Var}(\sigma^2\varepsilon^2\Delta t) = E[\sigma^4\varepsilon^4(\Delta t)^2] - [E(\sigma^2\varepsilon^2\Delta t)]^2 = 2\sigma^4(\Delta t)^2$$

其中对于标准正态随机变量，我们利用了 $E(\varepsilon^4)=3$. 这两个性质证明了 $\sigma^2\varepsilon^2\Delta t$ 当 $\Delta t\to 0$ 时收敛到一个非随机量 $\sigma^2\Delta t$. 因此，由方程(6.5)，我们有当 $\Delta t\to 0$ 时，

$$(\Delta x)^2 \to \sigma^2 \mathrm{d}t$$

将此结果插入方程(6.3)中，并且利用方程(6.2)中 x_t 的伊藤方程，我们得到

$$\begin{aligned}\mathrm{d}G &= \frac{\partial G}{\partial x}\mathrm{d}x + \frac{\partial G}{\partial t}\mathrm{d}t + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\sigma^2\mathrm{d}t \\ &= \left(\frac{\partial G}{\partial x}\mu + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\sigma^2\right)\mathrm{d}t + \frac{\partial G}{\partial x}\sigma\mathrm{d}w_t,\end{aligned}$$

这就是随机微积分中著名的伊藤引理.

注意我们在伊藤引理的推导中将变量(x_t, t)从漂移系数 μ 及波动率系数 σ 中省略了. 为了避免将来有任何的疑惑，我们将引理重述如下. [227]

伊藤引理

假定 x_t 是一个连续时间的随机过程，满足

$$\mathrm{d}x_t = \mu(x_t,t)\mathrm{d}t + \sigma(x_t,t)\mathrm{d}w_t$$

其中 w_t 是一个维纳过程. 另外，$G(x_t, t)$是 x_t 和 t 的可微函数. 那么

$$\mathrm{d}G = \left[\frac{\partial G}{\partial x}\mu(x_t,t) + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\sigma^2(x_t,t)\right]\mathrm{d}t + \frac{\partial G}{\partial x}\sigma(x_t,t)\mathrm{d}w_t \tag{6.6}$$

例 6.1 作为一个简单的解释，考虑维纳过程的平方函数 $G(w_t, t)=w_t^2$. 这里我们有 $\mu(w_t, t)=0$，$\sigma(w_t, t)=1$，而且

$$\frac{\partial G}{\partial w_t} = 2w_t,\quad \frac{\partial G}{\partial t} = 0,\quad \frac{\partial^2 G}{\partial w_t^2} = 2$$

因此，

$$\mathrm{d}w_t^2 = \left(2w_t \times 0 + 0 + \frac{1}{2}\times 2\times 1\right)\mathrm{d}t + 2w_t\mathrm{d}w_t = \mathrm{d}t + 2w_t\mathrm{d}w_t \tag{6.7}$$

6.3.3 一个应用

令 P_t 表示一只股票在时刻 t 的价格，它在$[0, \infty)$上连续. 文献中通常假定 P_t 服从一个特殊的伊藤过程

$$\mathrm{d}P_t = \mu P_t\mathrm{d}t + \sigma P_t\mathrm{d}w_t \tag{6.8}$$

其中 μ 与 σ 都是常数. 利用方程(6.2)中一般伊藤过程的记号，我们有 $\mu(x_t, t)=\mu x_t$，$\sigma(x_t, t)=\sigma x_t$，这里 $x_t=P_t$. 这样一个特殊的过程称为几何布朗运动. 现在我们利用伊藤引理来对股价 P_t 的对数得到一个连续时间模型. 令 $G(P_t, t)=\ln(P_t)$表示标的股票的对数价格，则我们有

$$\frac{\partial G}{\partial P_t} = \frac{1}{P_t},\quad \frac{\partial G}{\partial t} = 0,\quad \frac{1}{2}\frac{\partial^2 G}{\partial P_t^2} = \frac{1}{2}\frac{-1}{P_t^2}.$$

因此，根据伊藤引理，我们得到

$$\mathrm{d}\ln(P_t) = \left(\frac{1}{P_t}\mu P_t + \frac{1}{2}\frac{-1}{P_t^2}\sigma^2P_t^2\right)\mathrm{d}t + \frac{1}{P_t}\sigma P_t\mathrm{d}w_t = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma\mathrm{d}w_t$$

[228]

这个结果说明如果价格是几何布朗运动，那么价格的对数服从一个一般的维纳过程，其漂移率为 $\mu-\sigma^2/2$，方差率为 σ^2. 因此，价格的对数(即对数收益率) 从当前时刻 t 到将来某时刻 T 的变化服从均值为$(\mu-\sigma^2/2)(T-t)$，方差为 $\sigma^2(T-t)$的正态分布. 若时间间隔 $T-t=\Delta$ 是固定的，而且我们更感兴趣于对数价格的等间隔的增量，那么增量序列是一个均值为$(\mu-\sigma^2/2)\Delta$，方差为 $\sigma^2\Delta$ 的高斯过程.

6.3.4 μ 和 σ 的估计

方程(6.8)的几何布朗运动中两个未知参数 μ 与 σ 可以用经验方法来估计. 假定我们有股价 P_t 在等时间间隔 Δ(例如日、周或者月)上的 $n+1$ 个观测值，可用年来度量 Δ. 观测到的价格表示为$\{P_0, P_1, \cdots, P_n\}$，并且对 $t=1, \cdots, n$，令 $r_t=\ln(P_t)-\ln(P_{t-1})$.

因为 $P_t=P_{t-1}\exp(r_t)$，其中 r_t 是第 t 个时间间隔上的连续复合收益率. 利用前一小节中的结果，并且假定股价 P_t 服从一个几何布朗运动，我们得到 r_t 服从均值为$(\mu-\sigma^2/2)\Delta$，方差为 $\sigma^2\Delta$ 的正态分布. 另外，r_t 是前后不相关的.

为了简便，定义 $\mu_r=E(r_t)=(\mu-\sigma^2/2)\Delta$，$\sigma_r^2=\mathrm{Var}(r_t)=\sigma^2\Delta$. 令$\bar{r}$，$s_r$ 分别表示数据的样本均值和标准差，即

$$\bar{r}=\frac{\sum_{t=1}^{n} r_t}{n},\quad s_r=\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(r_t-\bar{r})^2}$$

正如第 1 章中提到的，$\bar{r}$与 s_r 分别是 r_i 的均值和标准差的相合估计，即当 $n\to\infty$时，$\bar{r}\to\mu_r$ 且 $s_r\to\sigma_r$. 因此，我们可以估计 σ，通过

$$\hat{\sigma}=\frac{s_r}{\sqrt{\Delta}}$$

另外，可以证明这个估计的标准差渐近于 $\hat{\sigma}/\sqrt{2n}$. 由 $\hat{\mu}_r=\bar{r}$，我们可以通过下式来估计 μ

$$\hat{\mu}=\frac{\bar{r}}{\Delta}+\frac{\hat{\sigma}^2}{2}=\frac{\bar{r}}{\Delta}+\frac{s_r^2}{2\Delta}$$

当序列 r_t 是前后相关的，或者当资产价格不服从方程(6.8)中的几何布朗运动时，则必须运用其他的估计方法来估计扩散方程的漂移参数和波动率参数. 稍后我们再讨论这个问题.

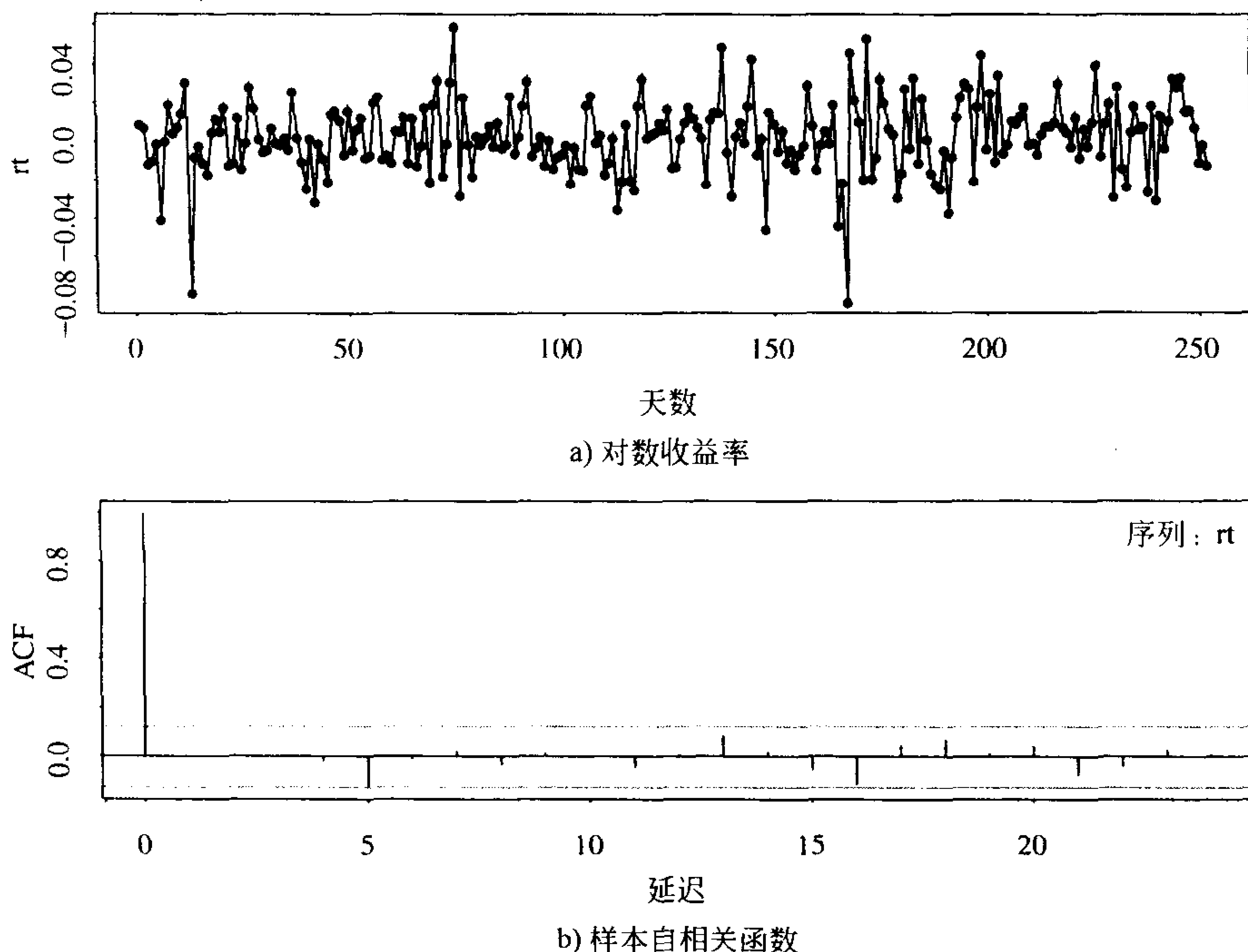
229

例 6.2 考虑 1998 年 IBM 股票的日对数收益率. 图 6-2a 是数据的时间图，有 252 个观测值. 图 6-2b 描述了序列的样本自相关. 可以看出，对数收益率确实是前后不相关的. L-B 统计量给出了 $Q(10)=4.9$，与自由度为 10 的 χ^2 分布比较是高度不显著的.

如果我们假定 1998 年 IBM 股票的价格服从方程(6.8)中的几何布朗运动，那么我们可以利用日对数收益率来估计参数 μ 和 σ. 由数据我们有$\bar{r}=0.002\,276$，$s_r=0.019\,15$. 因为一个交易日等价于 $\Delta=1/252$ 年，我们得到

$$\hat{\sigma}=\frac{s_r}{\sqrt{\Delta}}=0.304\,0,\quad \hat{\mu}=\frac{\bar{r}}{\Delta}+\frac{\hat{\sigma}^2}{2}=0.619\,8.$$

这样在 1998 年，IBM 股票每年的期望收益率的估计为 61.98%，标准差为 30.4%.



a) 对数收益率

b) 样本自相关函数

图 6-2　1998 年 IBM 股票的日收益率

然而，日对数收益率的正态性假定可能并不成立．在这个特殊的例子中，偏度－0.464(0.153)和超出峰度 2.396(0.306)增加了某种关注，这里括号内的数表示渐近标准误差． [230]

例 6.3　考虑 1999 年 Cisco Systems 股票的日对数收益率，共有 252 个观测值，样本均值和标准差分别为 0.003 32 和 0.026 303．对数收益率序列也显示没有前后相关性，$Q(12)=10.8$ 甚至在 10%水平下都是不显著的．所以，我们有

$$\hat{\sigma}=\frac{s_r}{\sqrt{\Delta}}=\frac{0.026\,303}{\sqrt{1.0/252.0}}=0.418,\quad \hat{\mu}=\frac{\bar{r}}{\Delta}+\frac{\hat{\sigma}^2}{2}=0.924$$

因此，1999 年 Cisco Systems 股票的年期望收益率的估计为 92.4%，标准差的估计为 41.8%．

6.4　股票价格与对数收益率的分布

前一节的结果显示，如果假定一支股票的价格服从几何布朗运动

$$\mathrm{d}P_t=\mu P_t\mathrm{d}t+\sigma P_t\mathrm{d}w_t$$

那么价格的对数服从一般的维纳过程

$$\mathrm{d}\ln(P_t)=\left(\mu-\frac{\sigma^2}{2}\right)\mathrm{d}t+\sigma\mathrm{d}w_t$$

其中 P_t 为股票在 t 时刻的价格，w_t 是一个维纳过程. 因此，对数价格从时刻 t 到时刻 T 的变化是正态分布的，为

$$\ln(P_T)-\ln(P_t)\sim N\left[\left(\mu-\frac{\sigma^2}{2}\right)(T-t),\sigma^2(T-t)\right] \tag{6.9}$$

因此，在给定时刻 t 的价格 P_t 的条件下，$T>t$ 时刻的对数价格是正态分布的，为

$$\ln(P_T)\sim N\left[\ln(P_t)+\left(\mu-\frac{\sigma^2}{2}\right)(T-t),\sigma^2(T-t)\right] \tag{6.10}$$

利用第 1 章中讨论的对数正态分布的结果，我们得到 P_T 的(条件)均值和方差分别为

$$E(P_T)=P_t\exp[\mu(T-t)]$$
$$\mathrm{Var}(P_T)=P_t^2\exp[2\mu(T-t)]\{\exp[\sigma^2(T-t)]-1\}$$

231 注意这个期望确认了 μ 是股票的期望收益率.

股价的先验分布可以用来作统计推断. 例如，假定股票 A 的当前价格为 50 美元，股票的年期望收益率为 15%，年波动率为 40%，则股票 A 在六个月(0.5 年)内的期望价格和相应的方差由下式给出

$$E(P_T)=50\exp(0.15\times0.5)=53.89$$
$$\mathrm{Var}(P_T)=2\,500\exp(0.3\times0.5)[\exp(0.16\times0.5)-1]=241.92$$

从现在开始的六个月内价格的标准差为$\sqrt{241.92}=15.55$.

下面，令 r 表示从时刻 t 到时刻 T 中每年的连续复合收益率，则我们有

$$P_T=P_t\exp[r(T-t)]$$

其中 T 和 t 是以年度量的. 因此，

$$r=\frac{1}{T-t}\ln\left(\frac{P_T}{P_t}\right)$$

由方程(6.9)，我们有

$$\ln\left(\frac{P_T}{P_t}\right)\sim N\left[\left(\mu-\frac{\sigma^2}{2}\right)(T-t),\sigma^2(T-t)\right].$$

因此，每年的连续复合收益率的分布为

$$r\sim N\left(\mu-\frac{\sigma^2}{2},\frac{\sigma^2}{T-t}\right).$$

所以，连续复合收益率服从均值为 $\mu-\sigma^2/2$，标准差为 $\sigma/\sqrt{T-t}$的正态分布.

考虑年期望收益率为 15%，年波动率为 10%的一支股票. 该股票在两年内的连续复合收益率的分布是正态的，每年的均值为 $0.15-0.01/2=0.145$ 或 14.5%，标准差为 $0.1/\sqrt{2}=0.071$ 或 7.1%. 这些结果允许我们对 r 构造置信区间(C.I.). 例如，r 的每年一个 95%的 C.I. 为 $0.145\pm1.96\times0.071$(即 0.6%，28.4%).

6.5 Black-Scholes 微分方程的推导

本节中，我们利用伊藤引理和无套利假定来推导价值为 P_t 的股票的衍生未定权益的价格的 B-S(Black-Scholes 的简称)微分方程．假定价格 P_t 服从方程(6.8)中的几何布朗运动，并且 $G_t=G(P_t, t)$是关于 P_t 的衍生未定权益(例如一个看涨期权)的价格．由伊藤引理 232

$$dG_t = \left(\frac{\partial G_t}{\partial P_t}\mu P_t + \frac{\partial G_t}{\partial t} + \frac{1}{2}\frac{\partial^2 G_t}{\partial P_t^2}\sigma^2 P_t^2\right)dt + \frac{\partial G_t}{\partial P_t}\sigma P_t dw_t$$

这个过程以及前面结果的离散化形式为

$$\Delta P_t = \mu P_t \Delta t + \sigma P_t \Delta w_t \tag{6.11}$$

$$\Delta G_t = \left(\frac{\partial G_t}{\partial P_t}\mu P_t + \frac{\partial G_t}{\partial t} + \frac{1}{2}\frac{\partial^2 G_t}{\partial P_t^2}\sigma^2 P_t^2\right)\Delta t + \frac{\partial G_t}{\partial P_t}\sigma P_t \Delta w_t \tag{6.12}$$

其中 ΔP_t，ΔG_t 分别为 P_t 和 G_t 在一个小时间区间 Δt 上的变化．因为对方程(6.11)和方程(6.12)都有 $\Delta w_t=\epsilon\sqrt{\Delta t}$，所以可以构造不涉及维纳过程的股票与衍生资产的一个组合．恰当的组合策略是卖空衍生资产并多头持有$\frac{\partial G_t}{\partial P_t}$的股份，用 V_t 表示组合的价值．由构造

$$V_t = -G_t + \frac{\partial G_t}{\partial P_t}P_t \tag{6.13}$$

则 V_t 的变化为

$$\Delta V_t = -\Delta G_t + \frac{\partial G_t}{\partial P_t}\Delta P_t \tag{6.14}$$

将方程(6.11)和方程(6.12)代入方程(6.14)，我们有

$$\Delta V_t = \left(-\frac{\partial G_t}{\partial t} - \frac{1}{2}\frac{\partial^2 G_t}{\partial P_t^2}\sigma^2 P_t^2\right)\Delta t \tag{6.15}$$

这个方程并不涉及随机部分 Δw_t．因此，在无套利假设下，组合 V_t 在一个小时间区间 Δt 上一定是无风险的．换句话说，所用的假设蕴涵组合一定同时可以赚得与其他的短期无风险证券相同的收益；否则，在这个资产组合与短期无风险证券之间就存在套利机会．因此，我们有

$$\Delta V_t = rV_t\Delta t \tag{6.16}$$

其中 r 为无风险利率．由方程(6.13)到方程(6.16)，我们有

$$\left(\frac{\partial G_t}{\partial t} + \frac{1}{2}\frac{\partial^2 G_t}{\partial P_t^2}\sigma^2 P_t^2\right)\Delta t = r\left(G_t - \frac{\partial G_t}{\partial P_t}P_t\right)\Delta t$$

所以 233

$$\frac{\partial G_t}{\partial t} + rP_t\frac{\partial G_t}{\partial P_t} + \frac{1}{2}\sigma^2 P_t^2\frac{\partial^2 G_t}{\partial P_t^2} = rG_t \tag{6.17}$$

这就是对衍生资产定价的 B-S 微分方程．可以通过解此方程来得到标的变量的衍生资产的价格 P_t，这样得到的解依赖于衍生资产的边界条件．对一个欧式看涨期权而言，边界条件为

$$G_T = \max(P_T - K, 0)$$

其中 T 是到期时间，K 是敲定价格．对一个欧式看跌期权而言，边界条件变为

$$G_T = \max(K - P_T, 0).$$

例 6.4 作为一个简单例子，考虑股票的一个远期合约，不支付红利．在这种情况下，合约的价值由下式给出

$$G_t = P_t - K\exp[-r(T-t)]$$

其中 K 是交割价格(delivery price)，r 是无风险利率，T 为到期日．对这样的一个函数，我们有

$$\frac{\partial G_t}{\partial t} = -rK\exp[-r(T-t)], \quad \frac{\partial G_t}{\partial P_t} = 1, \quad \frac{\partial^2 G_t}{\partial P_t^2} = 0$$

将这些量代入方程(6.17)的左面得到

$$-rK\exp[-r(T-t)] + rP_t = r\{P_t - K\exp[-r(T-t)]\}$$

等于方程(6.17)的右面．满足 B-S 微分方程．

6.6 Black-Scholes 定价公式

Black 与 Scholes(1973)成功地求解了方程(6.17)中的微分方程，并得到了对欧式看涨期权与看跌期权价格的精确公式．下面我们利用金融中所谓的无风险定价来推导这些公式．

6.6.1 风险中性世界

234 将漂移参数 μ 从 B-S 微分方程中去掉，金融上，这意味着此方程是与风险偏好独立的．换句话说，风险偏好不能影响这个方程的解．这个性质的一个完美结果是能够假设投资者是风险中性的．在一个风险中性世界里，我们有如下结论：

- 所有证券的期望收益率都是无风险利率 r，而且
- 任何现金流的当前价值可以通过将它的期望价值以无风险利率折现得到．

6.6.2 公式

在风险中性世界里，一个欧式看涨期权在到期日的期望价值为

$$E_*[\max(P_T - K, 0)]$$

其中 E_* 表示在无风险世界中的期望价值．看涨期权在 t 时刻的价格为

$$c_t = \exp[-r(T-t)]E_*[\max(P_T - K, 0)] \tag{6.18}$$

然而，在风险中性世界里，我们有 $\mu = r$，并且由方程(6.10)，$\ln(P_T)$是正态分布的：

$$\ln(P_T) \sim N\left[\ln(P_t) + \left(r - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right].$$

令 $g(P_T)$表示 P_T 的概率密度函数，则方程(6.18)中看涨期权的价格为

$$c_t = \exp[-r(T-t)]\int_K^\infty (P_T - K)g(P_T)\mathrm{d}P_T.$$

通过积分的变量变换以及一些代数计算(附录 A 中给出详细推导)，我们有

$$c_t = P_t\Phi(h_+) - K\exp[-r(T-t)]\Phi(h_-) \tag{6.19}$$

其中 $\Phi(x)$是标准正态随机变量的积累分布函数(CDF)在 x 点的值，

$$h_+ = \frac{\ln(P_t/K) + (r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},$$

$$h_- = \frac{\ln(P_t/K) + (r-\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} = h_+ - \sigma\sqrt{T-t}.$$

实际中，$\Phi(x)$可以通过大多数统计软件包很容易地得到. 另外一种可供选择的方法，可以运用附录 B 中给出的一个近似.

235

方程(6.19)中的 B-S 看涨期权公式有一些好的解释. 首先，如果我们在到期日执行期权，得到了股票，但是必须要支付敲定价格. 这个交换只有当期权是赚钱的(即 $P_T>K$)时才会发生. 当且仅当 $P_T>K$ 时，第一项 $P_t\Phi(h_+)$是得到股票的当前价值；当且仅当 $P_T>K$ 时，第二项$-K\exp[-r(T-t)]\Phi(h_-)$是支付敲定价格的当前价值. 第二个解释尤其有用，正如 6.5 节中 B-S 微分方程的推导中显示的，$\Phi(h_+)=\frac{\partial G_t}{\partial P_t}$是不涉及维纳过程的、资产组合中股份的数量. 我们知道 $c_t=P_t\Phi(h_+)+B_t$，其中 B_t 为资产组合(或衍生资产头寸)中无风险债券包含的美元总量. 可见从对 B-S 公式的检查中可以直接看出 $B_t=-K\exp[-r(T-t)]\Phi(h_-)$，公式的第一项 $P_t\Phi(h_+)$为投资在股票上的总量，而第二项 $K\exp[-r(T-t)]\Phi(h_-)$是借入的总量.

类似地，我们得到一个欧式看跌期权的价格为

$$p_t = K\exp[-r(T-t)]\Phi(-h_-) - P_t\Phi(-h_+). \tag{6.20}$$

因为标准正态分布是关于它的均值 0.0 对称的，所以我们有：对任何 x，$\Phi(x)=1-\Phi(-x)$. 利用这个性质，我们有 $\Phi(-h_i)=1-\Phi(h_i)$. 这样，计算一个看跌期权价格需要的信息与计算看涨期权价格所需要的信息是相同的. 另外一个方法，利用正态分布的对称性，很容易证明

$$p_t - c_t = K\exp[-r(T-t)] - P_t$$

这称为涨-跌平价，而且可用来从 c_t 中得到 p_t.

例 6.5 假设 Intel 股票的当前价格是每股 80 美元，年波动率为 $\sigma=20\%$，进一步假设年无风险利率为 8%. 那么敲定价格为 90 美元，而且在 3 个月内到期的 Intel 的欧式看涨期权的价格是多少?

由假设，我们有 $P_t=80$，$K=90$，$T-t=0.25$，$\sigma=0.2$，且 $r=0.08$. 因此，

$$h_+ = \frac{\ln(80/90) + (0.08+0.04/2)\times 0.25}{0.2\sqrt{0.25}} = -0.9278$$

$$h_- = h_+ - 0.2\sqrt{0.25} = -1.0278$$

利用统计软件(如 Minitab 或 SCA)，或者附录 B 中的近似式，我们有

$$\Phi(-0.9278) = 0.1767, \quad \Phi(-1.0278) = 0.1520$$

因此，一个欧式看涨期权的价格为

$$c_t = 80\Phi(-0.9278) - 90\Phi(-1.0278)\exp(-0.02) = 0.73(\text{美元})$$

236

对看涨期权的买者而言，只有股价升高 10.73 美元时，才能达到得失平衡.

在相同的假设下，欧式看跌期权的价格为

$$p_t = 90\exp(-0.08 \times 0.25)\Phi(1.027\,8) - 80\Phi(0.927\,8) = 8.95(\text{美元})$$

这样，对看跌期权的买者而言，股价可以升高额外的 1.05 美元而达到得失平衡.

例 6.6 前面例子中的敲定价格大大超出了当前股价. 一个更现实的敲定价格是 85 美元. 假设前面例子中其他的条件仍然成立，现在我们有 $P_t=80$，$K=85$，$r=0.08$，$T-t=0.25$，且 h_i 变为

$$h_+ = \frac{\ln(80/85) + (0.08 + 0.04/2) \times 0.25}{0.2\sqrt{0.25}} = -0.356\,246,$$

$$h_- = h_+ - 0.2\sqrt{0.25} = -0.456\,246.$$

利用附录 B 中的近似，我们有 $\Phi(-0.356\,246)=0.360\,8$ 和 $\Phi(-0.456\,246)=0.324\,1$，则一个欧式看涨期权的价格是

$$c_t = 80\Phi(-0.356\,246) - 85\exp(-0.02)\Phi(-0.456\,246) = 1.86(\text{美元})$$

对于看涨期权的买者而言，股价必须提高 6.86 美元，才得失相等. 然而，在同样假定下的一个欧式看跌期权的价格为

$$\begin{aligned} p_t &= 85\exp(-0.02)\Phi(0.456\,246) - 80\Phi(0.356\,246) \\ &= 85\exp(-0.02) \times 0.675\,9 - 80 \times 0.639\,2 = 5.18(\text{美元}) \end{aligned}$$

对看跌期权的买者而言，股票价格必须降低 0.18 美元，才得失相等.

6.6.3 讨论

由方程知，一个看涨或看跌期权的价格依赖于 5 个变量，即当前的股价 P_t，敲定价格 K，以年度量的到期日 $T-t$，年波动率 σ 以及年利率 r. 很值得研究的是这 5 个变量对期权价格的影响.

6.6.3.1 边际效应

首先考虑这 5 个变量对一个看涨期权价格 c_t 的边际效应. 边际效应的意思是在固定其余变量的情况下改变其中一个变量. 一个看涨期权的效用可以概括如下：

237

1) 当前股票价格 P_t：c_t 与 $\ln(P_t)$ 正相关. 特别地，当 $P_t \to 0$ 时，$c_t \to 0$；当 $P_t \to \infty$ 时，$c_t \to \infty$. 图 6-3a 解释了 $K=80$，年利率 $r=6\%$；$T-t=0.25$ 年以及年波动率 $\sigma=30\%$ 时的效应.

2) 敲定价格 K：c_t 与 $\ln(K)$ 负相关. 具体地，当 $K \to 0$ 时，$c_t \to P_t$；当 $K \to \infty$ 时，$c_t \to 0$.

3) 到期时间：c_t 与 $T-t$ 的相关性非常复杂，但是通过将 h_+ 和 h_- 写成

$$h_+ = \frac{\ln(P_t/K)}{\sigma\sqrt{T-t}} + \frac{(r+\sigma^2/2)\sqrt{T-t}}{\sigma}$$

$$h_- = \frac{\ln(P_t/K)}{\sigma\sqrt{T-t}} + \frac{(r-\sigma^2/2)\sqrt{T-t}}{\sigma}$$

可得到极限结果. 若 $P_t<K$，则当 $(T-t)\to 0$ 时 $c_t \to 0$. 若 $P_t>K$，则当 $(T-t)\to 0$ 时 $c_t \to P_t - K$，且当 $(T-t)\to\infty$ 时 $c_t \to P_t$. 图 6-4a 显示的是对三种不同的当前股价来说，$T-t$ 对 c_t 的边

际效应. 固定的变量是 $K=80$, $r=6\%$ 和 $\sigma=30\%$. 实线、点划线以及虚线分别对应与 $P_t=70$, 80 和 90.

4)波动率 σ: 通过将 h_+ 和 h_- 改写成

$$h_+=\frac{\ln(P_t/K)+r(T-t)}{\sigma\sqrt{T-t}}+\frac{\sigma}{2}\sqrt{T-t}$$

$$h_-=\frac{\ln(P_t/K)+r(T-t)}{\sigma\sqrt{T-t}}-\frac{\sigma}{2}\sqrt{T-t}$$

我们得到

(a)如果 $\ln(P_t/K)+r(T-t)<0$, 则当 $\sigma\to 0$ 时, $c_t\to 0$.

(b)如果 $\ln(P_t/K)+r(T-t)\geqslant 0$, 则当 $\sigma\to 0$ 时, $c_t\to P_t-Ke^{-r(T-t)}$; 当 $\sigma\to\infty$时, $c_t\to P_t$. 图 6-5a 显示了 $K=80$, $T-t=0.25$, $r=0.06$ 以及 P_t 取三个不同值时 σ 对 c_t 的效应. 实线、点划线以及虚线分别对应于 $P_t=70$, 80 和 90.

5)利率: c_t 与 r 是正相关的, 满足: 当 $r\to\infty$时, $c_t\to P_t$.

这 5 个变量对一个看跌期权的边际效应可以类似得到. 图 6-3b、图 6-4b、图 6-5b 对所选择的一些情形解释了其效应.

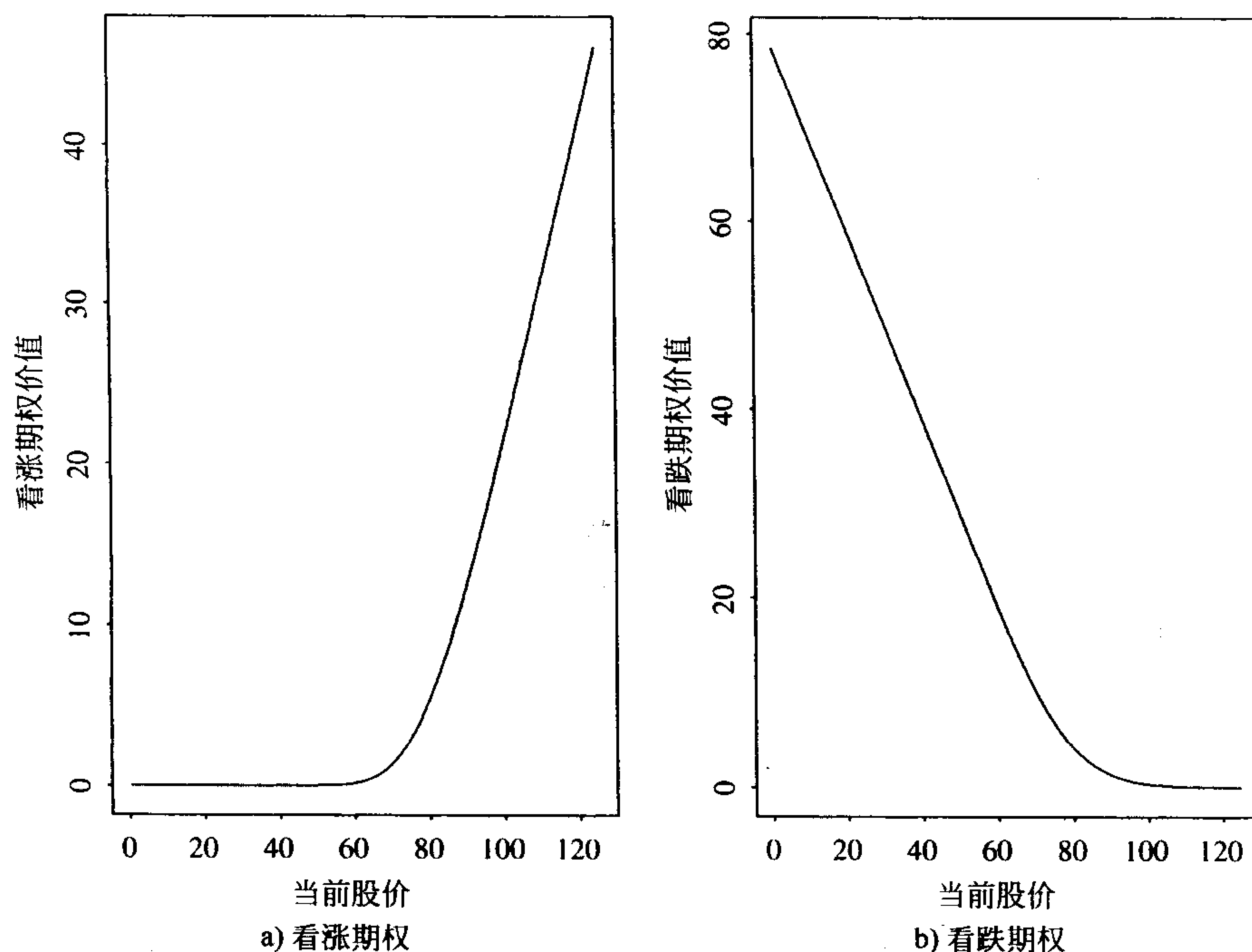
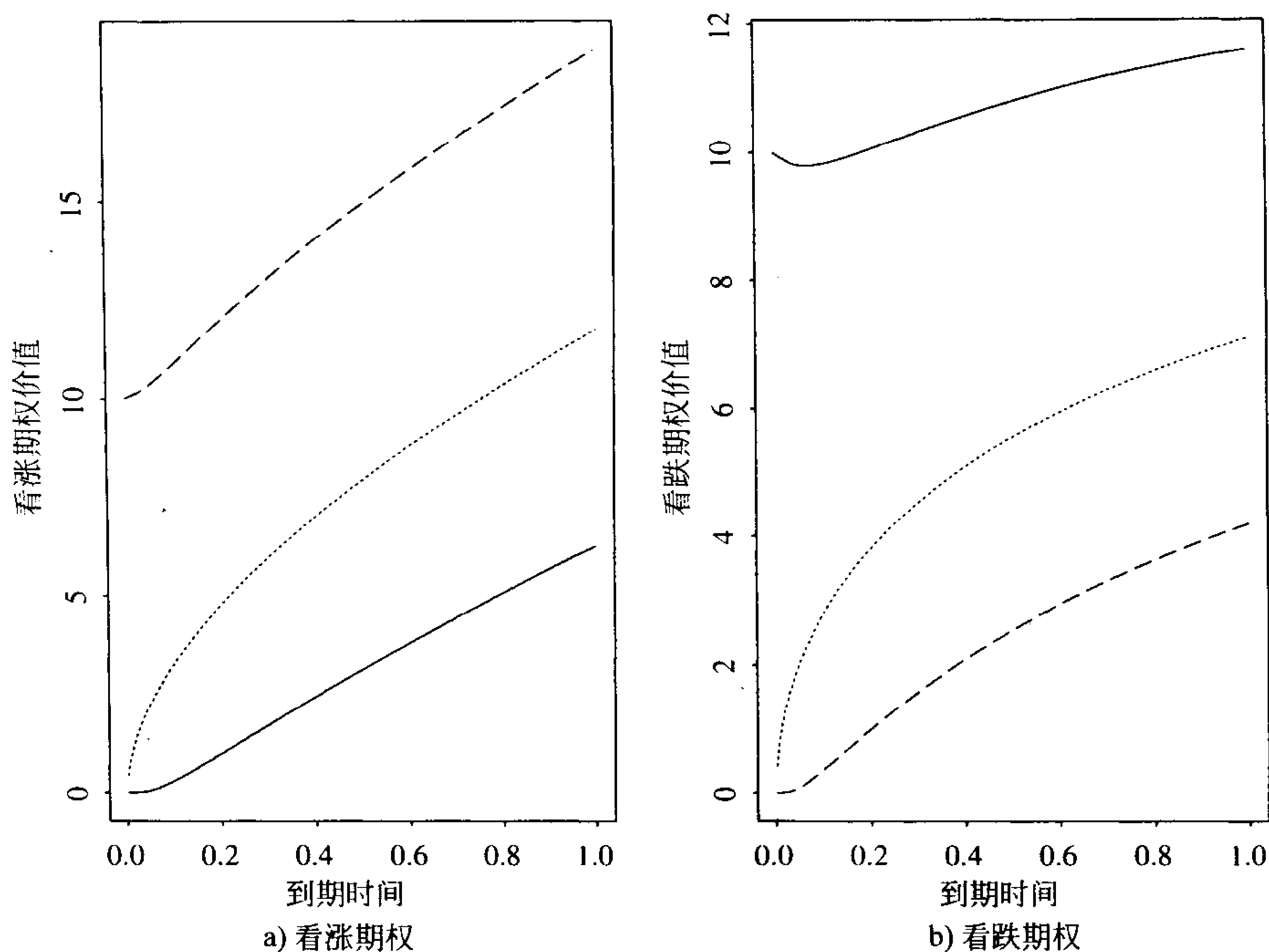


图 6-3 当前股价对期权价格的边际效应, 其中 $K=80$, $T-t=0.25$, $\sigma=0.3$, $r=0.06$



238～239

图 6-4 到期时间对期权价格的边际效应，其中 $K=80$，$T-t=0.25$，$\sigma=0.3$，$r=0.06$
实线、点划线和虚线分别表示当前的股价 $P_t=70$，80 和 90

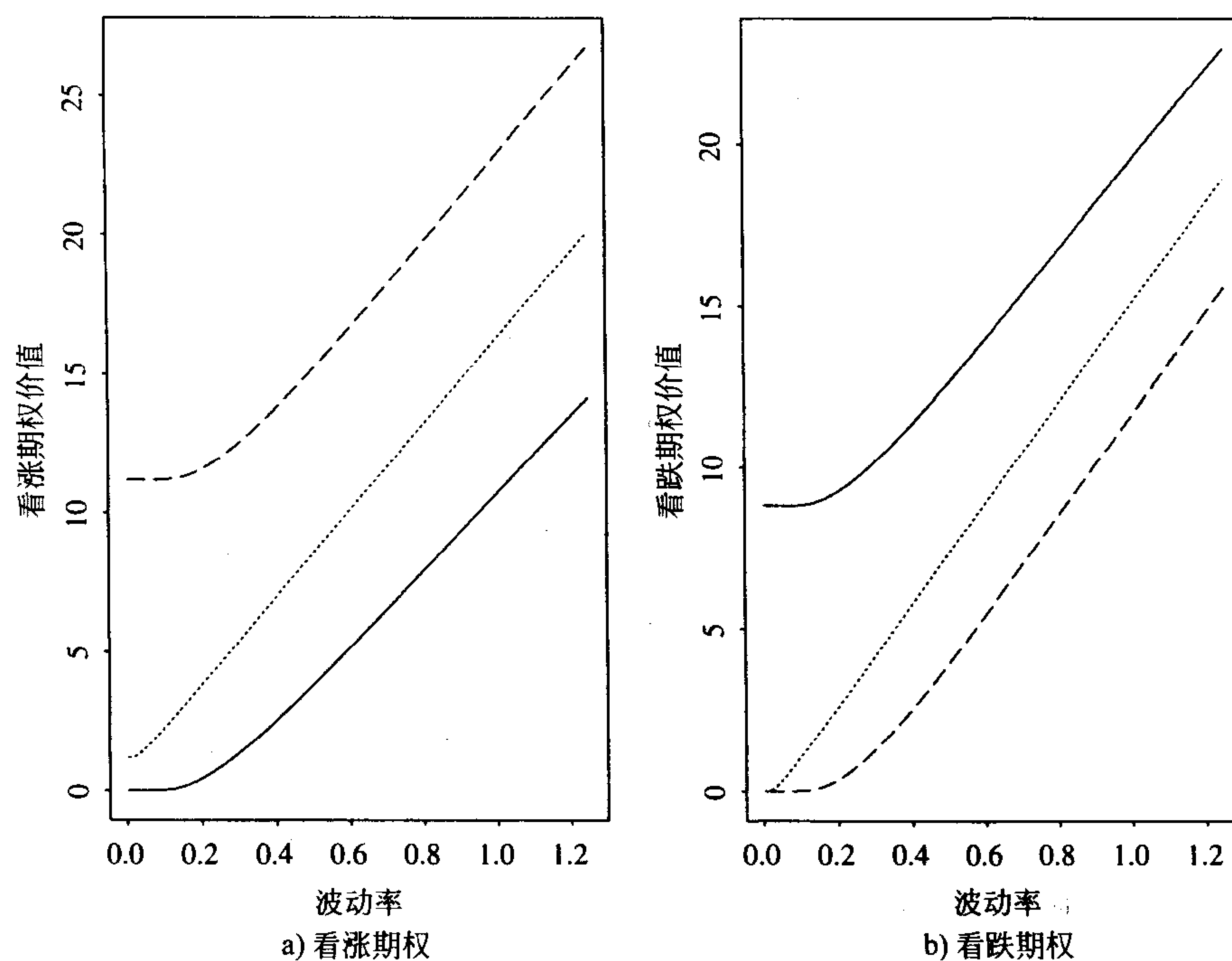


图 6-5 股市波动率对期权价格的边际效应，其中 $K=80$，$T-t=0.25$，$r=0.06$
实线、点划线和虚线分别表示当前的股价 $P_t=70$，80 和 90

6.6.3.2 一些联合效应

图 6-6 显示波动率与敲定价格对一个看涨期权的联合效应，这里其余变量是固定的，$P_t=80$，$r=0.06$，$T-t=0.25$. 正如所预料的，当波动率很高而且敲定价格正好低于当前的股价时，看涨期权的价格更高. 图 6-7 显示同样条件下，对一个看跌期权的效应. 当波动率很高而

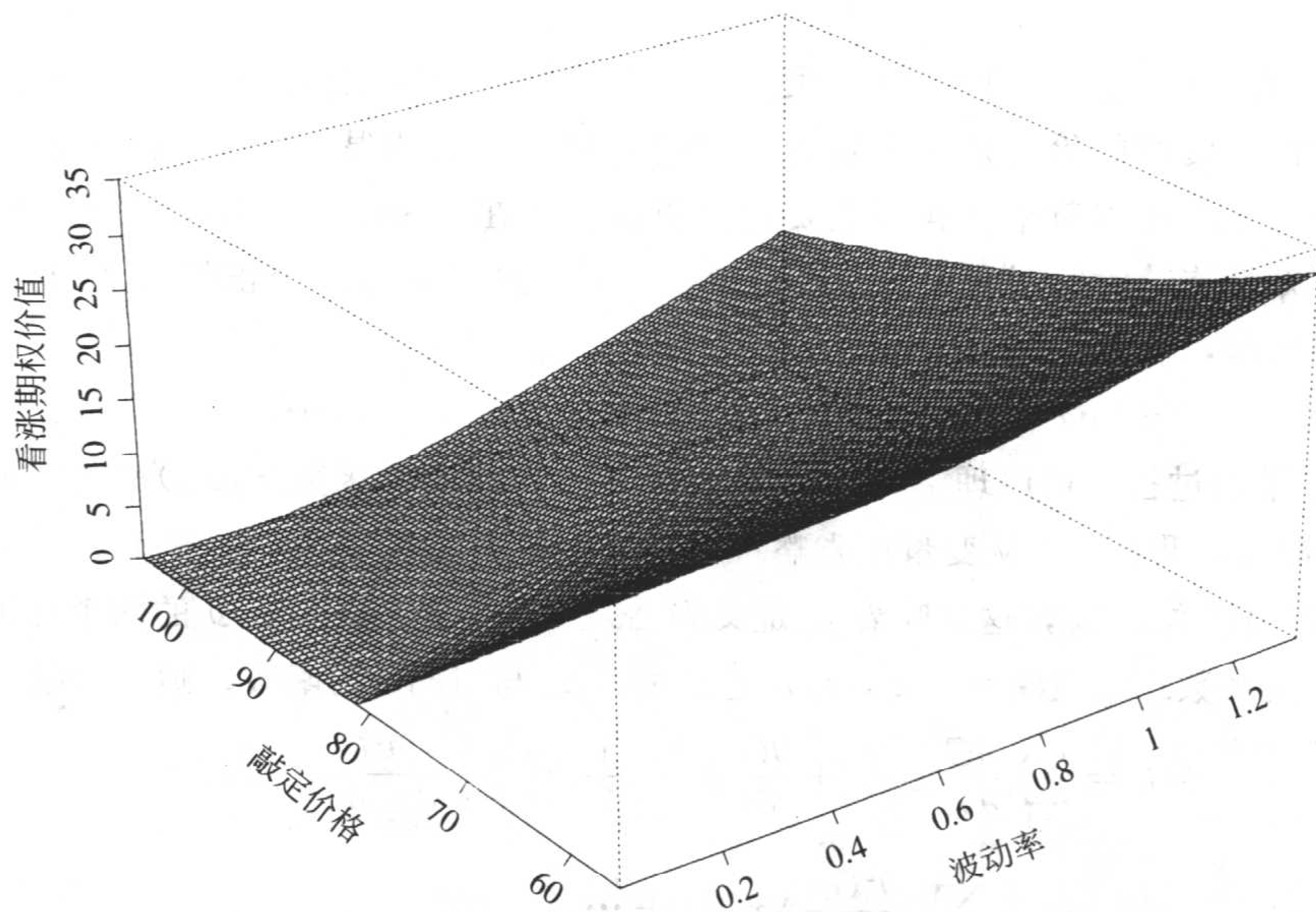


图 6-6 股市波动率与敲定价格对看涨期权的联合效应，其中 $P_t=80$，$T-t=0.25$，$r=0.06$

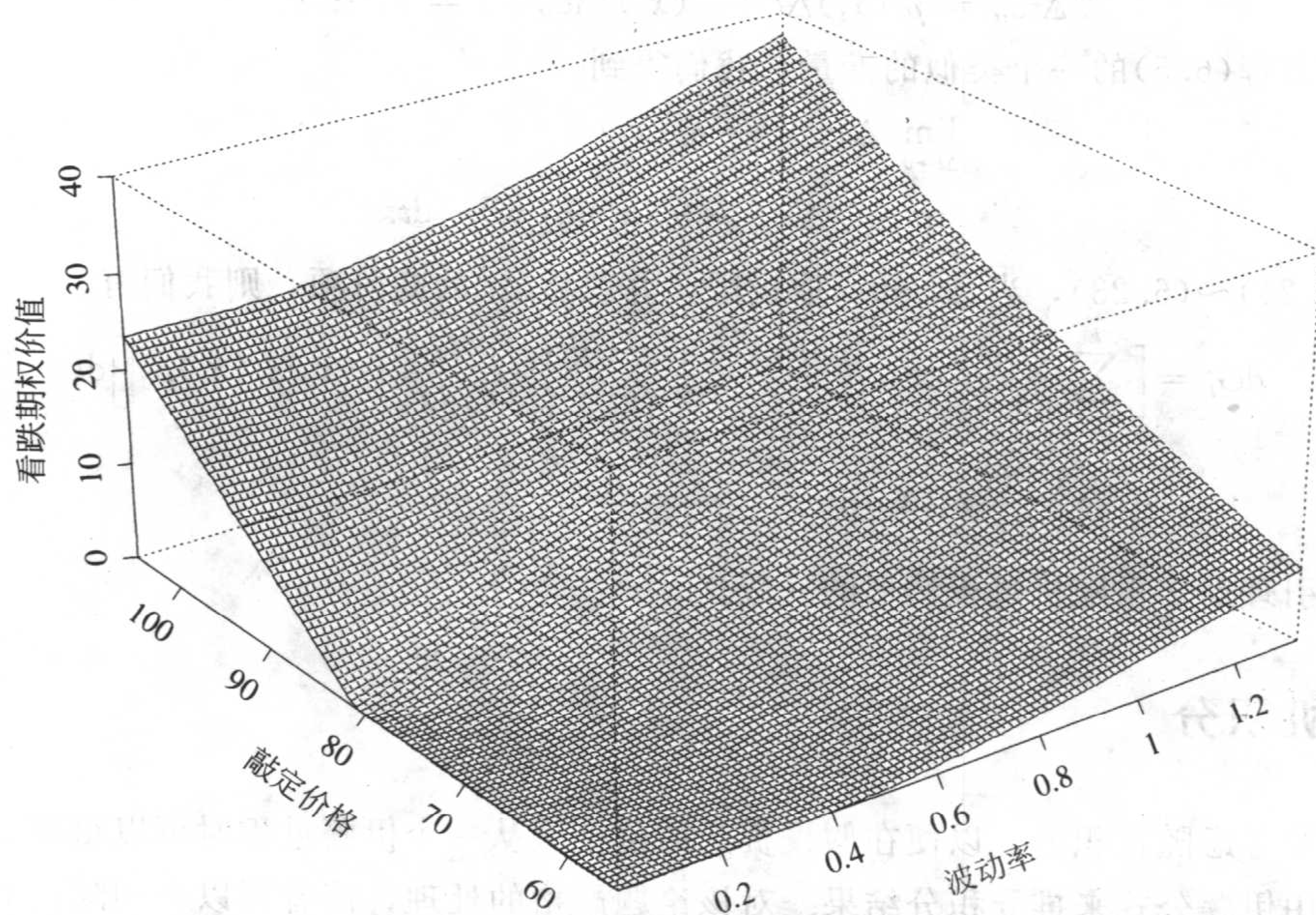


图 6-7 股市波动率与敲定价格对看跌期权的联合效应，其中 $K=80$，$T-t=0.25$，$r=0.06$

且敲定价格正好高于当前的股价时，看跌期权的价格更高. 而且，图形也说明了随着波动率的增加，敲定价格对看跌期权价格的效应变得更加线性化.

6.7 伊藤引理的扩展

240 ~ 241

在推导定价公式时，一个衍生资产可能是多种证券的未定权益. 当这些证券的价格是由多因素驱动时，衍生资产的价格是几个随机过程的函数. 利率期限结构的两因子模型就是二维随机过程的一个例子. 在本节中，我们主要讨论伊藤引理在几种随机过程情形下的扩展.

考虑一个 k 维连续时间过程 $\boldsymbol{x}_t=(x_{1t},\cdots,x_{kt})'$. 这里 k 是一个正整数，而且 x_{it} 是一个连续时间的随机过程，满足

$$dx_{it}=\mu_i(\boldsymbol{x}_t)dt+\sigma_i(\boldsymbol{x}_t)dw_{it},\qquad i=1,\cdots,k \tag{6.21}$$

其中 w_{it} 是一个维纳过程. 可以理解为漂移函数 $\mu_i(x_{it})$ 与波动率函数 $\sigma_i(x_{it})$ 也是时间指数 t 的函数. 为了简化记号，我们将 t 从变量中省略. 对 $i\neq j$，维纳过程 w_{it} 和 w_{jt} 是不一样的. 我们假定 dw_{it} 与 dw_{jt} 的相关系数是 ρ_{ij}，这意味着 ρ_{ij} 定义为 $\Delta w_{it}=\epsilon_i\Delta t$ 和 $\Delta w_{jt}=\epsilon_j\Delta t$ 的两个标准正态随机变量 ϵ_i 和 ϵ_j 的相关系数. 假设$G_t=G(\boldsymbol{x}_t,t)$是随机过程 x_{it} 与时间 t 的函数，则由泰勒展开得

$$\Delta G_t=\sum_{i=1}^{k}\frac{\partial G_t}{\partial x_{it}}\Delta x_{it}+\frac{\partial G_t}{\partial t}\Delta t+\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2 G_t}{\partial x_{it}\partial x_{jt}}\Delta x_{it}\Delta x_{jt}$$
$$+\frac{1}{2}\sum_{i=1}^{k}\frac{\partial^2 G_t}{\partial x_{it}\partial t}\Delta x_{it}\Delta t+\cdots$$

方程(6.21)的离散形式为

$$\Delta w_{it}=\mu_i(\boldsymbol{x}_t)\Delta t+\sigma_i(\boldsymbol{x}_t)\Delta w_{it}\quad i=1,\cdots,k.$$

利用 6.3 节方程(6.5)的一个类似的变量，我们得到

$$\lim_{\Delta t\to 0}(\Delta x_{it})^2\to\sigma_i^2(\boldsymbol{x}_t)dt \tag{6.22}$$

$$\lim_{\Delta t\to 0}\Delta x_{it}\Delta x_{jt}\to\sigma_i(\boldsymbol{x}_t)\sigma_j(\boldsymbol{x}_t)\rho_{ij}dt. \tag{6.23}$$

利用方程(6.21)～(6.23)，当 $\Delta t\to 0$，取极限，并忽略 Δt 的高阶项，则我们有

$$dG_t=\left[\sum_{i=1}^{k}\frac{\partial G_t}{\partial x_{it}}\mu_i(\boldsymbol{x}_t)+\frac{\partial G_t}{\partial t}+\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2 G_t}{\partial x_{it}\partial x_{jt}}\sigma_i(\boldsymbol{x}_t)\sigma_j(\boldsymbol{x}_t)\rho_{ij}\right]dt$$
$$+\sum_{i=1}^{k}\frac{\partial G_t}{\partial x_{it}}\sigma_i(\boldsymbol{x}_t)dw_{it} \tag{6.24}$$

这就是伊藤引理对多重随机过程情形的一般化.

6.8 随机积分

我们主要考虑随机积分，以便在假设资产的价格服从一个伊藤过程时可以得到该资产的价格. 我们利用伊藤公式来推导积分结果. 对该论题严格的处理，读者可以查阅随机微积分的教科书. 首先，如同一个确定函数的通常积分一样，积分是微分的反面，因此，对一个随机过程

242

x_t，仍然成立

$$\int_0^t dx_s = x_t - x_0.$$

具体地，对维纳过程 w_t，因为 $w_0=0$，所以我们有 $\int_0^t dw_s = w_t$. 下面，考虑积分 $\int_0^t w_s dw_s$. 利用前面的结果，并在方程(6.7)中取积分，我们有

$$w_t^2 = t + 2\int_0^t w_s dw_s,$$

因此，

$$\int_0^t w_s dw_s = \frac{1}{2}(w_t^2 - t).$$

这不同于通常的确定积分，那里

$$\int_0^t y dy = (y_t^2 - y_0^2)/2.$$

转到 x_t 是一个几何布朗运动的情况，即 x_t 满足

$$dx_t = \mu x_t dt + \sigma x_t dw_t,$$

其中 μ，σ 是常数，$\sigma>0$；见方程(6.8). 对 $G(x_t, t)=\ln(x_t)$ 利用伊藤引理，我们得到

$$d\ln(x_t) = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dw_t.$$

取积分，并利用前面得到的结果，我们有

$$\int_0^t d\ln(x_s) = \left(\mu - \frac{\sigma^2}{2}\right)\int_0^t ds + \sigma\int_0^t dw_s.$$

因此，

$$\ln(x_t) = \ln(x_0) + (\mu - \sigma^2/2)t + \sigma w_t$$

而且

$$x_t = x_0 \exp[(\mu - \sigma^2/2)t + \sigma w_t].$$

243

对资产价格将记号 x_t 改为 P_t，则我们在假设价格服从几何布朗运动时，价格有一个解，此价格为

$$P_t = P_0 \exp[(\mu - \sigma^2/2)t + \sigma w_t]. \tag{6.25}$$

6.9 跳跃扩散模型

经验研究发现，基于布朗运动的随机扩散模型不能解释资产收益率与它们的衍生资产价格的一些特征(例如，隐含波动率的"波动率微笑"，见 Bakshi，Cao 和 Chen(1997)及参考文献). 波动率微笑指的是期权的隐含波动率与敲定价格的一个凸函数. 赔钱和赚钱的期权都倾向于比不赔不赚的期权有更高的隐含波动率，尤其是在外汇市场中. 证券期权对波动率微笑的重视较少. 标准随机扩散模型的不足导致了其他连续时间模型的发展. 例如，文献中已提出跳跃扩散

及随机波动率模型克服来这种不足；见 Merton(1976) 和 Duffie(1995).

股价的跳跃经常假定为服从一个概率法则. 例如，跳跃可能服从一个泊松过程，它是一个连续时间的离散过程. 对给定的时刻 t，令 X_t 表示在时间段[0，t]上一个特定事件发生的次数. 如果

$$Pr(X_t = m) = \frac{\lambda^m t^m}{m!}\exp(-\lambda t), \qquad \lambda > 0,$$

则 X_t 是一个泊松过程，也就是，X_t 服从一个参数为 λt 的泊松分布. 参数 λ 控制了特定事件的发生，称为过程的强度. 正式的定义还要求 X_t 是一个带有左极限的右连续的齐次马尔可夫过程.

本节中，我们讨论 Kou(2000) 提出的一个简单的跳跃扩散模型，这个简单模型有几个好的性质. 模型隐含的收益率是有尖峰的、关于 0 点非对称的. 另外，模型可以再生波动率微笑并对许多期权的价格提供解析公式. 模型包括两部分，第一部分是连续的，服从几何布朗运动；第二部分是一个跳跃过程. 跳跃的产生是由一个泊松过程控制的，跳跃的大小服从一个双指数分布. 令 P_t 表示资产在 t 时刻的价格，简单的跳跃扩散模型假定价格服从随机微分方程

244

$$\frac{dP_t}{P_t} = \mu dt + \sigma dw_t + d\Big(\sum_{i=1}^{n_t}(J_i - 1)\Big) \tag{6.26}$$

其中 w_t 是一个维纳过程，n_t 是强度为 λ 的泊松过程，$\{J_i\}$是独立同分布的非负随机变量序列，满足 $X=\ln(J)$服从双重指数分布，而且概率密度函数为

$$f_X(x) = \frac{1}{2\eta}e^{-|x-\kappa|/\eta}, \qquad 0 < \eta < 1. \tag{6.27}$$

在模型(6.26)中，n_t、w_t 和 J_i 是相互独立的，以至模型的随机性之间没有关系，注意到 n_t 是时间间隔[0，t]上发生跳跃的次数，并服从参数 λt 的泊松分布，其中 λ 是一个常数. 在第 i 次跳跃中，价格跳跃的比例为 J_i-1.

双指数分布可写为

$$X - \kappa = \begin{cases} \xi & \text{以概率 } 0.5 \\ -\xi & \text{以概率 } 0.5 \end{cases} \tag{6.28}$$

其中 ξ 是均值为 η，方差为 η^2 的指数随机变量. ξ 的概率密度函数为

$$f(x) = \frac{1}{\eta}e^{-x/\eta}, \qquad 0 < x < \infty$$

双指数分布的一些有用性质为

$$E(X) = \kappa, \quad \mathrm{Var}(X) = 2\eta^2, \quad E(e^X) = \frac{e^\kappa}{1-\eta^2}.$$

对有限样本，很难区别双指数分布与学生-t 分布. 然而，双指数分布在解析上更易于处理，并且可以产生一个在均值周围的更高的概率集中度(例如更高峰). 如第 1 章中所述，观测到的资产收益率的直方图倾向于比正态密度有更高的峰. 图 6-8 用实线显示了双指数随机变量的概率

密度函数，虚线是一个正态随机变量的概率密度函数．两个变量的均值都是 0，方差为 0.000 8．可以很清楚看到双指数密度的高峰．

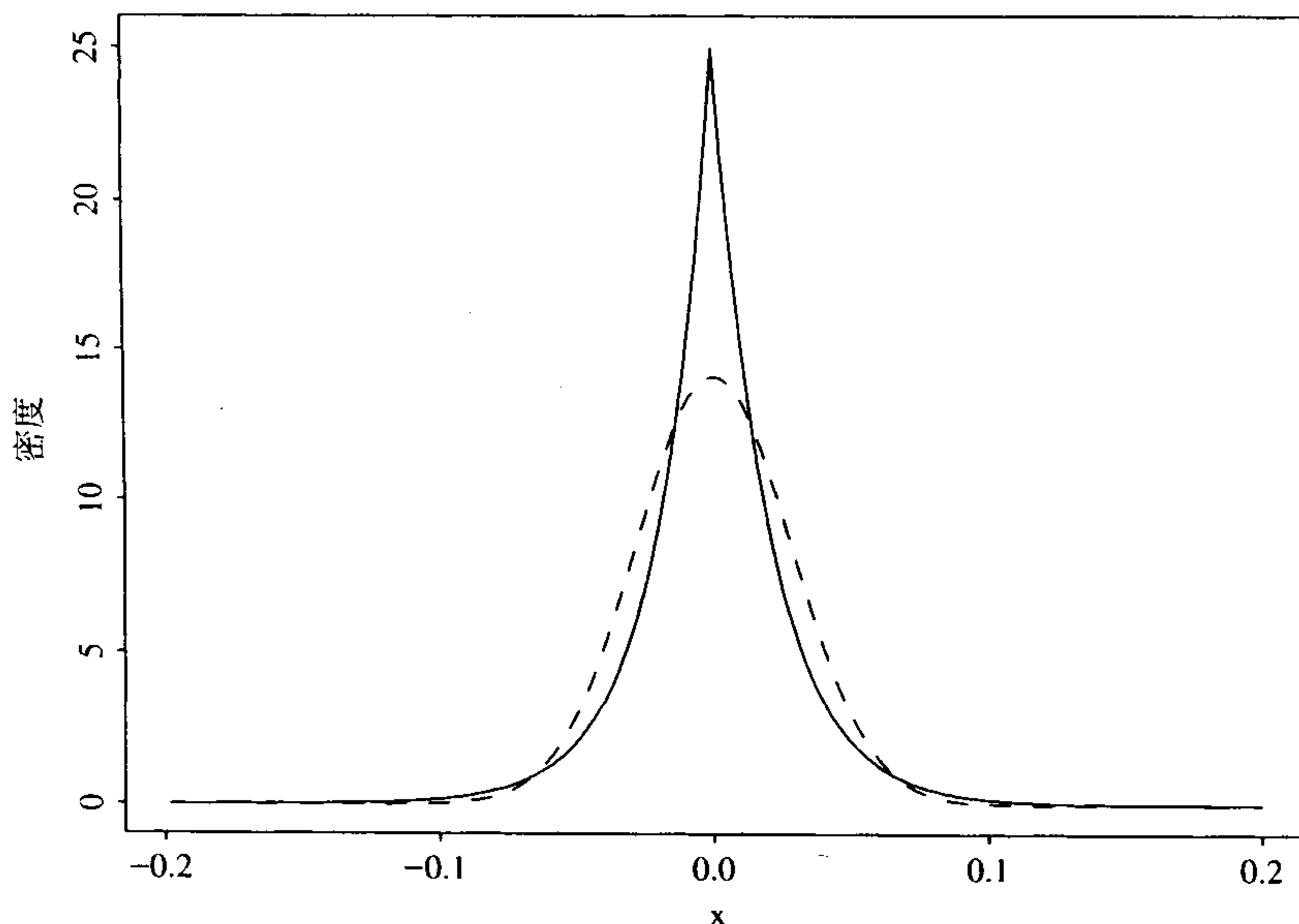


图 6-8　双指数及正态随机变量的概率密度函数，均值为零，方差为 0.000 8．实线表示双指数分布

解方程(6.26)的随机差分方程，我们得到资产价格的动态规律为

$$P_t = P_0 \exp[(\mu - \sigma^2/2)t + \sigma w_t] \prod_{i=1}^{n_t} J_i \tag{6.29}$$

这里 $\prod_{i=1}^{0} = 1$．这个结果是方程(6.25)包含随机跳跃的一般化，可以如下得到．令 t_i 表示第 i 次跳跃的时间，对 $t \in [0, t_1)$，没有跳跃而且价格可以由方程(6.25)给出．因此，左面价格在时刻 t_1 时极限为

$$P_{t_1^-} = P_0 \exp[(\mu - \sigma^2/2)t_1 + \sigma w_{t_1}].$$

在时刻 t_1，价格跳跃的比例为 $J_1 - 1$，以至价格变为

$$P_{t_1} = (1 + J_1 - 1)P_{t_1^-} = J_1 P_{t_1^-} = P_0 \exp[(\mu - \sigma^2/2)t_1 + \sigma w_{t_1}]J_1.$$

对 $t \in (t_1, t_2)$，区间$(t_1, t]$上没有跳跃，所以

$$P_t = P_{t_1} \exp[(\mu - \sigma^2/2)(t - t_1) + \sigma(w_t - w_{t_1})].$$

插入 P_{t_1} 中，我们有

$$P_t = P_0 \exp[(\mu - \sigma^2/2)t + \sigma w_t]J_1$$

重复这个方案，得到方程(6.29)．

由方程(6.29)，标的资产在一个小的时间增量 Δt 上的简单收益率为

$$\frac{P_{t+\Delta t} - P_t}{P_t} = \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma(w_{t+\Delta t} - w_t) + \sum_{i=n_t+1}^{n_{t+\Delta t}} X_i\right] - 1,$$

245～246

这里，一个空集上的和理解为 0，且 $X_i=\ln(J_i)$. 对一个小 Δt，我们可以利用近似

$$e^x \approx 1+x+x^2/2,$$

以及 6.3 节讨论的结果$(\Delta w_t)^2\approx\Delta t$，得到

$$\frac{P_{t+\Delta t}-P_t}{P_t}\approx\left(\mu-\frac{1}{2}\sigma^2\right)\Delta t+\sigma\Delta w_t+\sum_{i=n_t+1}^{n_{t+\Delta t}}X_i+\frac{1}{2}\sigma^2(\Delta w_t)^2$$

$$\approx\mu\Delta t+\sigma\epsilon\sqrt{\Delta t}+\sum_{i=n_t+1}^{n_{t+\Delta t}}X_i,$$

其中 $\Delta w_t=w_{t+\Delta t}-w_t$，且 ϵ 是一个标准正态随机变量.

在泊松过程的假设下，在时间间隔$(t, t+\Delta t]$上有一个跳跃的概率为 $\lambda\Delta t$，多于一个跳跃的概率为 $o(\Delta t)$，其中 $o(\Delta t)$意指：如果我们将这一项用 Δt 来除，则当 Δt 趋于 0 时，它的值也趋于 0. 因此，对一个小 Δt，忽略掉多个跳跃，我们有

$$\sum_{i=n_t+1}^{n_{t+\Delta t}}X_i\approx\begin{cases}X_{n_t+1} & \text{以概率 } \lambda\Delta t\\ 0 & \text{以概率 } 1-\lambda\Delta t\end{cases}$$

与前面的结果结合，可以看出标的资产的简单收益率的近似分布可由下式给出

$$\frac{P_{t+\Delta t}-P_t}{P_t}\approx\mu\Delta t+\sigma\epsilon\sqrt{\Delta t}+I\times X \tag{6.30}$$

其中 I 是伯努利(Bernoulli)随机变量，$Pr(I=1)=\lambda\Delta t$，$Pr(I=0)=1-\lambda\Delta t$；$X$ 是方程(6.28)中定义的双指数随机变量. 无跳跃时方程(6.30)简化为一个几何布朗运动.

令 $G=\mu\Delta t+\sigma\epsilon\sqrt{\Delta t}+I\times X$ 表示方程(6.30)右面的随机变量，利用模型中用到的指数分布与正态分布的独立性，Kou(2000) 得到了 G 的概率密度函数为

$$g(x)=\frac{\lambda\Delta t}{2\eta}e^{\sigma^2\Delta t/(2\eta^2)}\left\{e^{-\omega/\eta}\Phi\left(\frac{\omega\eta-\sigma^2\Delta t}{\sigma\eta\sqrt{\Delta t}}\right)+e^{\omega/\eta}\Phi\left(\frac{\omega\eta+\sigma^2\Delta t}{\sigma\eta\sqrt{\Delta t}}\right)\right\}$$
$$+(1-\lambda\Delta t)\frac{1}{\sigma\sqrt{\Delta t}}f\left(\frac{x-\mu\Delta t}{\sigma\sqrt{\Delta t}}\right) \tag{6.31}$$

其中 $\omega=x-\mu\Delta t-\kappa$，$f(.)$和 $\Phi(.)$分别是标准正态随机变量的概率密度函数和积累分布函数. 另外，

247

$$E(G)=\mu\Delta t+\kappa\lambda\Delta t,\quad \mathrm{Var}(G)=\sigma^2\Delta t+\lambda\Delta t[2\eta^2+\kappa^2(1-\lambda\Delta t)].$$

图 6-9 显示了正态分布和方程(6.31)的分布的概率密度函数之间的一些比较. 两个分布的均值都是 0，方差 2.0572×10^{-4}. 均值和方差是通过假定标的资产的收益率满足每年的 $\mu=20\%$，每年的 $\sigma=20\%$，$\Delta t=1$ 天$=1/252$ 年，$\lambda=10$，$\kappa=-0.02$，$\eta=0.02$得到的. 换句话说，我们假定每年大约有 10 天跳跃，平均跳跃大小为-2%，跳跃波动率为 2%，这些值对美国股票是合理的. 从图中，清楚地看到由方程(6.26)的跳跃扩散过程推导出的分布的尖峰态性质. 该分布比相应的正态分布有更高的峰和更厚的尾部.

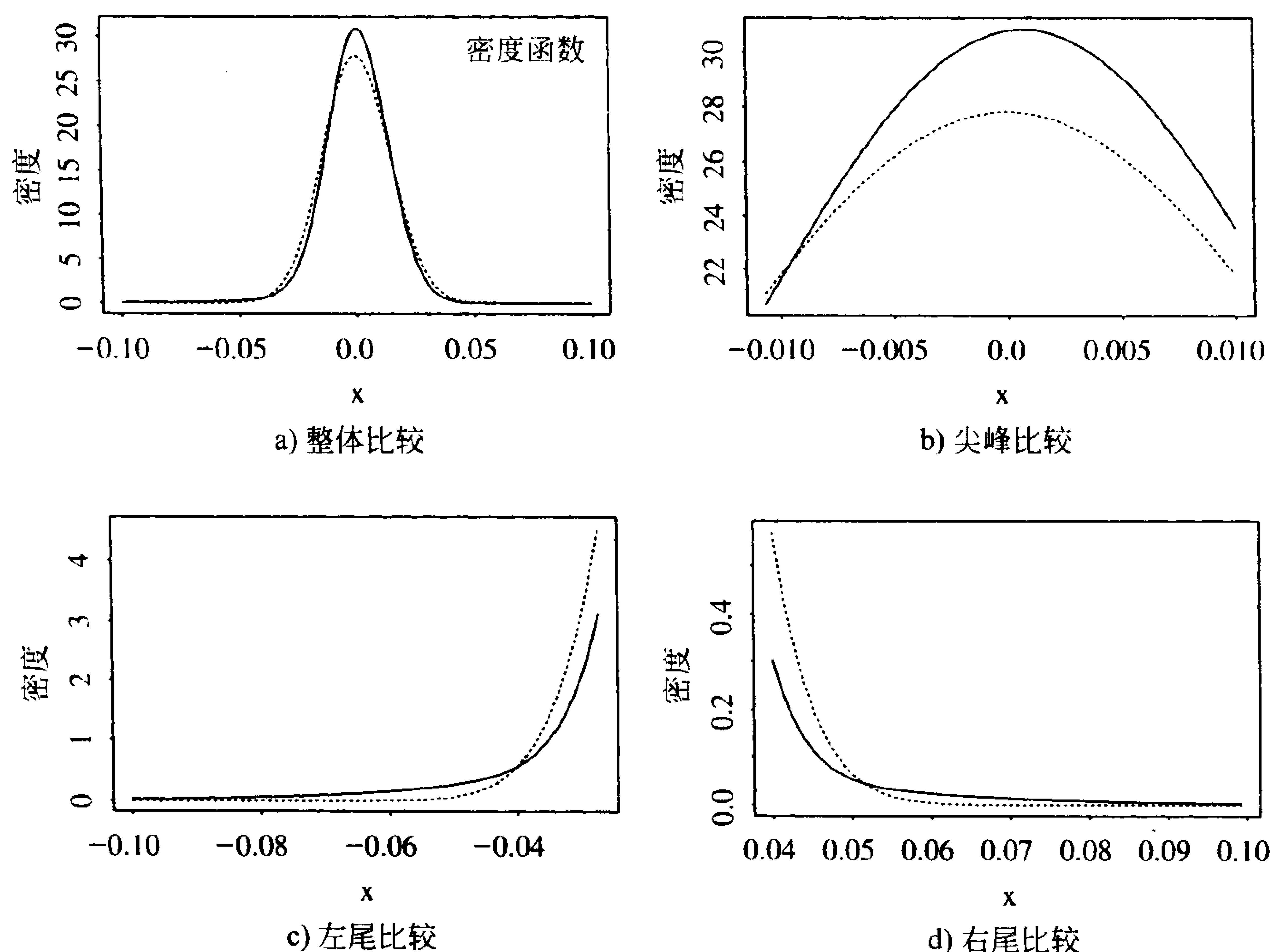


图 6-9　正态分布和式(6.31)的分布密度的比较．点线表示正态分布．两个分布的均值都是 0，方差 2.0572×10^{-4}．

跳跃扩散下的期权定价

当随机跳跃出现时，市场变成不完全的．在这种情况下，标准套期变量对期权定价就不是实际可行的．但我们仍可以推导出不依赖于对风险的态度的期权定价公式，通过假定可利用的证券数量非常大，以至突然跳跃的风险是可以分散的；而且市场对超过无风险利率的部分承担风险时不支付风险溢价．换一种说法，对风险溢价的一个给定集合，可以考虑风险中性测度 P^*，满足 248

$$\frac{\mathrm{d}P_t}{P_t}=[r-\lambda E(J-1)]\mathrm{d}t+\sigma\mathrm{d}w_t+\mathrm{d}\Big[\sum_{i=1}^{n_t}(J_i-1)\Big]$$

$$=(r-\lambda\psi)\mathrm{d}t+\sigma\mathrm{d}w_t+\mathrm{d}\Big[\sum_{i=1}^{n_t}(J_i-1)\Big]$$

其中 r 是无风险利率，$J=\exp(X)$满足 X 服从方程(6.27)的双指数分布，$\psi=\mathrm{e}^{\kappa}/(1-\eta^2)-1(0<\eta<1)$．考虑风险溢价时，参数 κ，η，ψ 和 σ 变成了风险中性参数，详见 Kou(2000)．前面方程的唯一解为

$$P_t=P_0\exp\Big[\Big(r-\frac{\sigma^2}{2}-\lambda\psi\Big)t+\sigma w_t\Big]\prod_{i=1}^{n_t}J_i.$$

为了对跳跃扩散模型中的一个欧式期权定价，剩下的是计算期权折现的最终回报在测度 P^* 下

的期望. 具体地，一个欧氏看涨期权在 t 时刻的价格给出为

$$
\begin{aligned}
c_t &= E_*[\mathrm{c}^{-r(T-t)}(P_T-K)_+] \\
&= E_*\left[\mathrm{e}^{-r(T-t)}\left(P_t\exp\left[\left(\frac{r-\sigma^2}{2-\lambda\psi}\right)(T-t)+\sigma\sqrt{T-t\varepsilon}\right]\prod_{i=1}^{n_T}J_i-K\right)_+\right]
\end{aligned} \tag{6.32}
$$

其中 T 是到期时间，$(T-t)$是用年测量的离到期的时间，K 是敲定价格，$(y)_+=\max(0,\ y)$，ε 是标准正态随机变量. Kou(2000)证明了 c_t 在解析上很容易处理，因为

$$
\begin{aligned}
c_t &= \sum_{n=1}^{\infty}\sum_{j=1}^{n}\mathrm{e}^{-\lambda(T-t)}\ \frac{\lambda^n(T-t)^n}{n!}\ \frac{2^j}{2^{2n-1}}\binom{2n-j-1}{n-1} \\
&\quad\times(A_{1,n,j}+A_{2,n,j}+A_{3,n,j})+\mathrm{e}^{-\lambda(T-t)}[P_t\mathrm{e}^{-\lambda\psi(T-t)}\Phi(h_+)-K\mathrm{e}^{-r(T-t)}\Phi(h_-)],
\end{aligned} \tag{6.33}
$$

其中 $\Phi(.)$是标准正态随机变量的积累分布函数，

[249]

$$
\begin{aligned}
A_{1,n,j} &= P_t\mathrm{e}^{-\lambda\psi(T-t)+n\kappa}\ \frac{1}{2}\left[\frac{1}{(1-\eta)^j}+\frac{1}{(1+\eta)^j}\right]\Phi(b_+)-\mathrm{e}^{-r(T-t)}K\Phi(b_-) \\
A_{2,n,j} &= \frac{1}{2}\mathrm{e}^{-r(T-t)-\omega/\eta+\sigma^2(T-t)/(2\eta^2)}K\times\sum_{i=0}^{j-1}\left[\frac{1}{(1-\eta)^{j-i}}-1\right]\left(\frac{\sigma\sqrt{T-t}}{\eta}\right)^i\frac{1}{\sqrt{2\pi}}Hh_i(c_-) \\
A_{3,n,j} &= \frac{1}{2}\mathrm{e}^{-r(T-t)+\omega/\eta+\sigma^2(T-t)/(2\eta^2)}K\times\sum_{i=0}^{j-1}\left[1-\frac{1}{(1+\eta)^{j-i}}\right]\left(\frac{\sigma\sqrt{T-t}}{\eta}\right)^i\frac{1}{\sqrt{2\pi}}Hh_i(c_+) \\
b_\pm &= \frac{\ln(P_t/K)+(r\pm\sigma^2/2-\lambda\psi)(T-t)+n\kappa}{\sigma\sqrt{T-t}} \\
h_\pm &= \frac{\ln(P_t/K)+(r\pm\sigma^2/2-\lambda\psi)(T-t)}{\sigma\sqrt{T-t}} \\
c_\pm &= \frac{\sigma\sqrt{T-t}}{\eta}\pm\frac{\omega}{\sigma\sqrt{T-t}} \\
\omega &= \ln(K/P_t)+\lambda\psi(T-t)-(r-\sigma^2/2)(T-t)-n\kappa \\
\psi &= \frac{\mathrm{e}^{\kappa}}{1-\eta^2}-1
\end{aligned}
$$

而函数 $Hh_i(.)$定义为

$$
Hh_n(x)=\frac{1}{n!}\int_x^{\infty}(s-x)^n\mathrm{e}^{-s^2/2}\mathrm{d}s,\qquad n=0,1,\cdots \tag{6.34}
$$

且有 $Hh_{-1}(x)=\exp(-x^2/2)$，当 $f(x)$是一个正态随机变量的概率密度函数时，它就是 $\sqrt{2\pi}f(x)$；见 Abramowitz 和 Stegun(1972). 函数 $Hh_n(x)$满足下面的递推公式

$$
nHh_n(x)=Hh_{n-2}(x)-xHh_{n-1}(x),\qquad n\geqslant 1 \tag{6.35}
$$

初始值为 $Hh_{-1}(x)=\mathrm{e}^{-x^2/2}$，且 $Hh_0(x)=\sqrt{2\pi}\Phi(-x)$.

定价公式涉及一个无穷序列，但是它的数值可以通过截断(truncation)(例如，前 10 项)快速而精确地逼近. 如果 $\lambda=0$(即没有跳跃)，则很容易看到对前面讨论的看涨期权而言，c_t 简化为 B-S 公式.

最后，考虑跳跃扩散模型的条件下，一个欧式看跌期权的价格可以通过涨-跌平价得

到，即

$$p_t = c_t + Ke^{-r(T-t)} - P_t.$$

250

在方程(6.26)的跳跃扩散模型下的其他期权定价公式可以参见 Kou(2000).

例 6.7 考虑例 6.6 中的股票，当前价格为 80 美元，如前，假定欧式期权的敲定价格为 $K=85$美元，其他参数为 $r=0.08$，$T-t=0.25$. 另外，假设股价服从方程(6.26)中的跳跃扩散模型，参数为 $\lambda=10$，$\kappa=-0.02$，$\eta=0.02$. 换句话说，每年大约有 10 个跳跃，平均跳跃大小为-2%，跳跃波动率为 2%. 利用方程(6.33)中的公式，我们得到 $c_t=2.25$ 美元，当没有跳跃时，大大高于例 6.6 中的 1.86 美元. 相应的看跌期权假定为 $p_t=5.57$ 美元，也比我们以前的高很多. 如所料想，当保持其余参数不变，加入跳跃时增加了两种欧式期权的价格. 然而，请注意，实际应用时，在股价中加入跳跃过程经常导致股市波动率 σ 的不同估计.

6.10 连续时间模型的估计

下面我们考虑直接从离散样本数据中估计扩散方程(即伊藤过程) 的问题. 这里漂移函数 $\mu(x_t, t)$和波动率函数 $\sigma(x_t, t)$都是时变的，可能不服从一个具体的参数形式. 这是近年来相当引起人们兴趣的一个主题. 具体方法的细节超出了本章的范围，因此，我们仅仅对文献中提出的方法简要介绍一下，有兴趣的读者可以查阅相应的文献以及 Lo(1988).

估计一个扩散方程时有几种可以利用的方法. 第一种方法是拟最大似然方法，它利用了 dw_t 在一个很小的时间间隔上是正态分布的事实，参见 Kessler(1997) 及其参考文献. 第二种方法是矩方法，参见 Conley，Hansen，Luttmer 和 Scheinkman(1997)及其参考文献. 第三种方法是利用非参数方法，见 Ait-Sahalia(1996，1997). 第四种方法是利用半参数方法和再投影(reprojection)方法，见 Gallant 和 Long(1997)，Gallant 和 Tauchen(1997). 最近，许多研究者已经利用 MCMC 方法来估计扩散方程，见 Eraker(2001)，Elerian，Chib 和 Shephard(2001).

附录 A B-S 公式积分

在本附录中，我们推导出方程(6.19) 给出的欧式看涨期权的价格. 令 $x=\ln(P_T)$，利用变量代换，以及 $g(P_T)dP_T=f(x)dx$，其中 $f(x)$为 x 的概率密度函数，我们有

251

$$\begin{aligned} c_t &= \exp[-r(T-t)]\int_K^{\infty}(P_T-K)g(P_T)dP_t \\ &= e^{-r(T-t)}\int_{\ln(K)}^{\infty}(e^x-K)f(x)dx \\ &= e^{-r(T-t)}\left[\int_{\ln(K)}^{\infty}e^x f(x)dx - K\int_{\ln(K)}^{\infty}f(x)dx\right] \end{aligned} \tag{6.36}$$

因为 $x=\ln(P_T)\sim N[\ln(P_t)+(r-\sigma^2/2)(T-t), \sigma^2(T-t)]$，所以方程(6.36)中第二项的积分简化为

$$\int_{\ln(K)}^{\infty}f(x)dx = 1-\int_{-\infty}^{\ln(K)}f(x)dx$$

$$= 1 - \mathrm{CDF}(\ln(K))$$
$$= 1 - \Phi(-h_-) = \Phi(h_-)$$

其中 $\mathrm{CDF}(\ln(K))$为 $x=\ln(P_T)$的积累分布函数在 $\ln(K)$处的取值，$\Phi(.)$是标准正态随机变量的积累分布函数，而且

$$-h_- = \frac{\ln(K) - \ln(P_t) - (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$
$$= \frac{-\ln(P_t/K) - (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}.$$

方程(6.36)中第一项的积分可以写为

$$\int_{\ln(K)}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(T-t)}} \exp\left[x - \frac{[x - \ln(P_t) - (r - \sigma^2/2)(T-t)]^2}{2\sigma^2(T-t)}\right] \mathrm{d}x,$$

其中指数可以简化为

$$x - \frac{[x - \ln(P_t) - (r - \sigma^2/2)(T-t)]^2}{2\sigma^2(T-t)}$$
$$= -\frac{[x - \{\ln(P_t) + (r + \sigma^2/2)(T-t)\}]^2}{2\sigma^2(T-t)} + \ln(P_t) + r(T-t).$$

因此，第一个积分变为

$$\int_{\ln(K)}^{\infty} \mathrm{e}^x f(x)\,\mathrm{d}x$$
$$= P_t \mathrm{e}^{r(T-t)} \int_{\ln(K)}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(T-t)}} \exp\left[-\frac{[x - \{\ln(P_t) + (r + \sigma^2/2)(T-t)\}]^2}{2\sigma^2(T-t)}\right] \mathrm{d}x$$

252

它包含了一个均值为 $\ln(P_t)+(r+\sigma^2/2)(T-t)$、方差为 $\sigma^2(T-t)$的正态分布的积累分布函数. 利用前面证明第二个积分的同样方法，我们有

$$\int_{\ln(K)}^{\infty} \mathrm{e}^x f(x)\,\mathrm{d}x = P_t \mathrm{e}^{r(T-t)} \Phi(h_+)$$

其中 h_+ 由下式给出

$$h_+ = \frac{\ln(P_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

将两个结果结合，我们就有

$$c_t = \mathrm{e}^{-r(T-t)}[P_t \mathrm{e}^{r(T-t)} \Phi(h_+) - K\Phi(h_-)] = P_t \Phi(h_+) - K\mathrm{e}^{-r(T-t)} \Phi(h_-).$$

附录 B 标准正态概率的近似

一个标准正态随机变量的积累分布函数 $\mathrm{CDF}\Phi(x)$可以近似为

$$\Phi(x) = \begin{cases} 1 - f(x)[c_1 k + c_2 k^2 + c_3 k^3 + c_4 k^4 + c_5 k^5] & \text{若 } x \geq 0 \\ 1 - \Phi(-x) & \text{若 } x < 0, \end{cases}$$

其中 $f(x)=\exp(-x^2/2)/\sqrt{2\pi}$，$k=1/(1+0.2316419x)$，$c_1=0.319381530$，$c_2=-0.356563782$，$c_3=1.781477937$，$c_4=-1.821255978$，$c_5=1.330274429$.

利用前面的近似，我们得到 $\Phi(1.96)=0.975002$，$\Phi(0.82)=0.793892$，$\Phi(-0.61)=0.270931$. 这些概率与一个典型的正态概率表中得到的值是非常接近的.

练习题

1. 假定对数价格 $p_t=\ln(P_t)$ 服从一个随机微分方程

$$dp_t=\gamma dt+\sigma dw_t$$

其中 w_t 是一个维纳过程. 试推导价格 P_t 的随机方程.

2. 考虑一个不支付红利的股票的远期价格 F，我们有

$$F_{t,T}=P_t e^{r(T-t)}$$

其中 r 是无风险利率，它是常数；P_t 是当前的股价. 假设 P_t 服从几何布朗运动

$$dP_t=\mu P_t dt+\sigma P_t dw_t$$

试推导 $F_{t,T}$ 的随机微分方程.

253

3. 假定 IBM 股票的价格服从伊藤过程

$$dP_t=\mu P_t dt+\sigma P_t dw_t,$$

其中 μ 和 σ 都是常数，w_t 是一个标准的布朗运动. 考虑 IBM 股票在 1997 年的日对数收益率. 平均收益率与样本的标准差分别为 0.001 31 和 0.022 15. 假设 1997 年有 252 个交易日，运用此数据来估计参数 μ 和 σ.

4. 假设一只股票的当前价格是每股 120 美元，年波动率 $\sigma=50\%$. 进一步假定年无风险利率为 7%，且该股票不支付红利. (a)这只股票的敲定价格为 125 美元，到期时间为 3 个月的欧式看涨期权的价格是什么？(b) 这只股票的敲定价格为 118 美元，到期时间为 3 个月的欧式看跌期权的价格是什么？如果年波动率 σ 增加到 80%，那么两种期权的价格是什么？

5. 推导一只股票的欧式看跌期权中五个变量 K，P_t，$T-t$，σ 及 r 的极限边际效应.

6. 股票当前的价格为每股 60 美元，并且服从几何布朗运动

$$dP_t=\mu P_t dt+\sigma P_t dt$$

假设该股票的年期望收益率 μ 为 20%，年波动率为 40%，问股价在两年内的概率分布是什么？试求该分布的均值和标准差，并且对此股价构造一个 95%的置信区间.

7. 当前股价为每股 60 美元，并服从几何布朗运动

$$dP_t=\mu P_t dt+\sigma P_t dt$$

假设该股票的年期望收益率 μ 为 20%，年波动率为 40%. 问股票两年的连续复合收益率的概率分布是什么？试求此分布的均值和标准差.

8. 假设股票 A 的当前价格是每股 70 美元，此价格服从方程(6.26)的跳跃扩散模型. 假定年无风险利率为 8%，年波动率为 30%. 另外，价格平均每年大约有 15 个跳跃，平均的跳跃大

小为−2%，跳跃波动率为 3%. 敲定价格为 75 美元，到期时间为 3 个月的一个欧式看涨期权的价格是多少？相应的欧式看跌期权的价格是多少？

参考文献

Abramowitz, M., and Stegun, I. A. (1972), *Handbook of Mathematical Functions*, 10th ed., U.S. National Bureau of Standards.

Ait-Sahalia, Y. (1996), "Testing continuous-time models for the spot interest rate," *Review of Financial Studies*, 9, 385–426.

254

Ait-Sahalia, Y. (1997), "Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach," working paper, Economics Department, Princeton University.

Bakshi, G., Cao, C., and Chen, Z. (1997), "Empirical performance of alternative option pricing models," *Journal of Finance*, 52, 2003–2049.

Billingsley, P. (1986), *Probability and Measure*, 2nd ed., Wiley: New York.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley: New York.

Black, F. and Scholes, M. (1973), "The pricing of options and corporate liabilities," *Journal of Political Economy*, 81, 637–654.

Conley, T. G., Hansen, L. P., Luttmer, E. G. J., and Scheinkman, J. A. (1997), "Short-term interest rates as subordinated diffusions," *Review of Financial Studies*, 10, 525–577.

Cox, J. C., and Rubinstein, M. (1985), *Options Markets*, Prentice-Hall: Englewood Cliffs, New Jersey.

Donsker, M. (1951), "An invariance principle for certain probability limit theorems," *Mem. American Mathematical Society*, No. 6.

Duffie, D. (1995), *Dynamic Asset Pricing Theory*, 2nd ed., Princeton University Press: Princeton, New Jersey.

Elerian, O., Chib, S., and Shephard, N. (2001), "Likelihood inference for discretely observed non-linear diffusions," *Econometrica*, 69, 959–993.

Eraker, B. (2001), "MCMC analysis of diffusion models with application to finance," *Journal of Business & Economic Statistics*, 19, 177–191.

Gallant, A. R., and Long, J. R. (1997), "Estimating stochastic diffusion equations efficiently by minimum chi-squared, " *Biometrika*, 84, 125–141.

Gallant, A. R., and Tauchen, G. (1997), "The relative efficiency of method of moments estimators," Working paper, Economics Department, University of North Carolina.

Hull, J. C. (1997), *Options, Futures, and Other Derivatives*, 3rd ed. Prentice-Hall: Upper Saddle River, New Jersey.

Kessler, M. (1997), "Estimation of an ergodic diffusion from discrete observations," *Scandinavian Journal of Statistics*, 24, 1–19.

Kou, S. (2000), "A jump diffusion model for option pricing with three properties: Leptokurtic feature, volatility smile, and analytic tractability," working paper, Columbia University.

Lo, A. W. (1988), "Maximum likelihood estimation of generalized Ito's processes with discretely sampled data," *Econometric Theory*, 4, 231–247.

255

Merton, R. C. (1976), "Option pricing when the underlying stock returns are discontinuous," *Journal of Financial Economics*, 5, 125–144.

第7章　极值理论、分位数估计与 VaR

金融市场中极端的价格运动虽然少见，但是很重要. 1987 年 10 月华尔街股市的崩溃以及其他大的金融危机(如长期资本管理，Long Term Capital Management)，已经引起了实际应用者和研究者们的广泛关注，一些人甚至呼吁政府加强对衍生证券市场的监管. 近年来，高科技股表面上很大的日价格变动进一步产生了对市场风险以及金融机构的边际设置的讨论. 因此，风险值(以下简记为 VaR) 成为风险管理中广泛使用的市场风险的度量.

本章中，我们讨论计算 VaR 的各种方法以及这些方法的统计理论. 我们特别考虑统计文献中发展起来的极值理论，研究稀少（或异常）事件，及其对 VaR 的应用. 极值理论的条件与无条件概念都在讨论之列. 金融头寸 VaR 计算的无条件方法用到计算 VaR 所涉及的金融工具的历史收益率. 然而，条件方法运用历史数据与解释变量来计算 VaR.

本章中讨论的其他计算 VaR 的方法是风险度量制（RiskMetrics)、波动率模型的经济计量模型、以及经验分位数. 我们利用 IBM 股票的日对数收益率来解释所有讨论方法的实际计算. 因此，得到的结果可以用来比较不同方法的表现. 图 7-1 描述了 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日的日对数收益率，共 9 190 个观测值的时间图.

7.1　VaR

金融市场中有几种类型的风险，如信用风险、流动性风险以及市场风险等. VaR 主要讨论市场风险，它是在一个风险范畴中的一个机构的头寸在一个给定持有期间内，由于一般的市场运动而降低，带来损失的统一估计. 对 VaR 的一般阐述，参见 Duffie 和 Pan (1997)，Jorion (1997). 金融机构可以利用这个度量评估风险，或者通过一个管理委员会设定边际要求. 在任何情形，都可以用 VaR 确保金融机构经过一个灾难性事件之后仍然可以运转. 从金融机构的角度，VaR 可以定义为金融头寸在一个给定时间段上，以一个给定的概率发生的最大损失. 以这种看法，可以将 VaR 看作是在正常市场条件下，与稀少（异常）事件相联系的损失的一个度量. 换种说法，从管理委员会的角度，VaR 可以定义为在异常市场情境下的最小损失. 两种定义尽管概念上貌似相差很远，但都导致同样的 VaR 度量.

256

下面，我们在概率框架中定义 VaR. 假设在时间指标 t 点，我们感兴趣的是接下来的 l 段中一个金融头寸的风险. 令 $\Delta V(l)$表示金融头寸中，从时刻 t 到时刻 $t+l$ 资产价值的变化，这个数量用美元度量，在时间指标为 t 时是一个随机变量. 用 $F_l(x)$表示 $\Delta V(l)$的积累分布函数(CDF)，定义一个多头头寸在持有期 l 中概率为 p 的 VaR 为

$$p = \Pr[\Delta V(l) \leqslant \text{VaR}] = F_l(\text{VaR}) \tag{7.1}$$

当 $\Delta V(l)<0$ 时，一个多头金融头寸的持有者遭受损失，当 p 很小时，方程(7.1)中定义的

VaR 是一个负值，负号表示一个损失. 由定义，在时间区间 l 上，持有者将要遭受的损失大于或等于 VaR 的概率为 p. 换一种说法，VaR 可以如下解释：以概率 $1-p$，金融头寸的持有者在时间区间 l 上遭受的损失小于或等于 VaR.

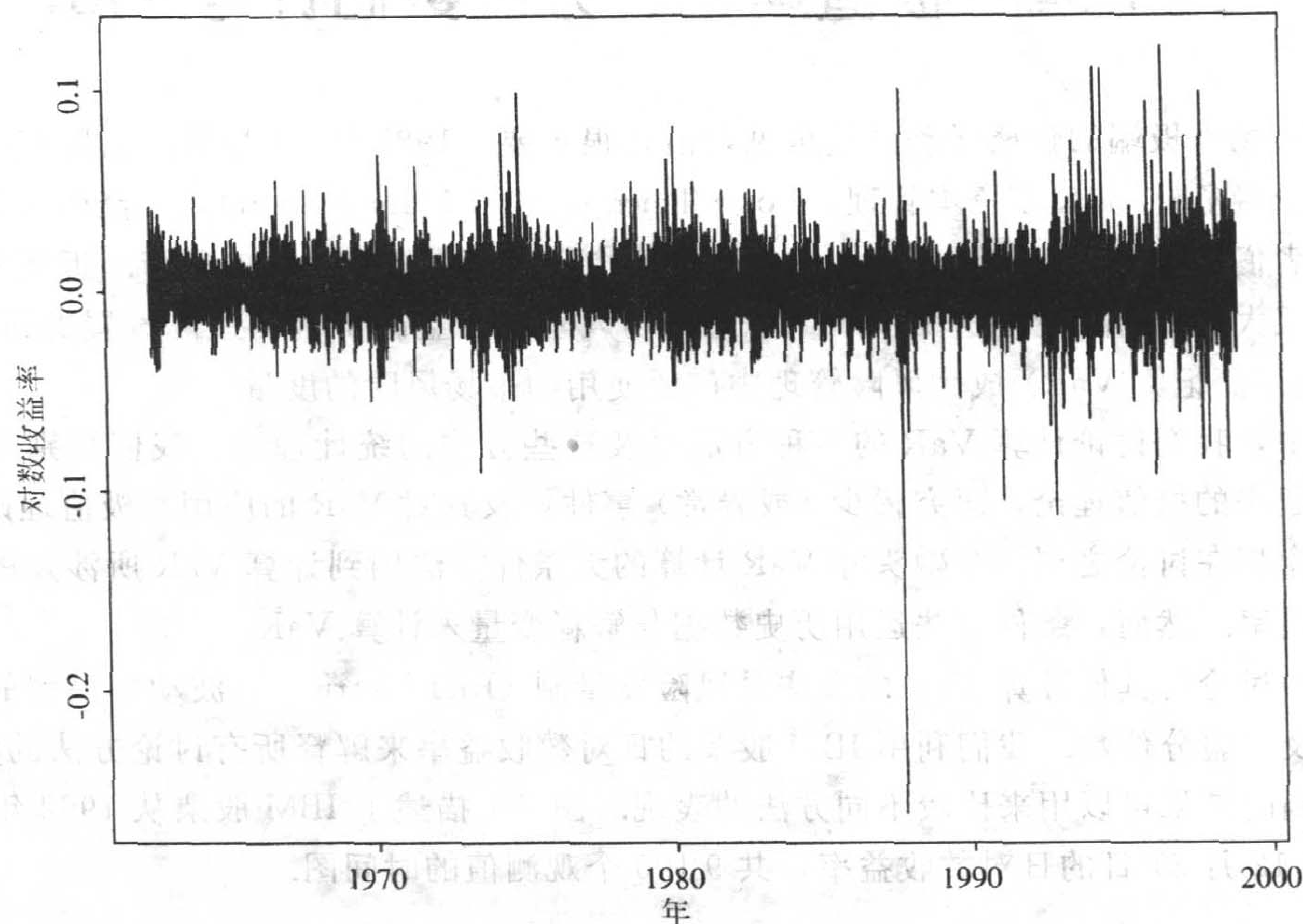


图 7-1 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日的日对数收益率的时间图

当资产价值增加(即 $\Delta V(l)>0$)时，空头头寸的持有者遭受损失，这时 VaR 定义为

$$p=\Pr[\Delta V(l)\geqslant \mathrm{VaR}]=1-\Pr[\Delta V(l)\leqslant \mathrm{VaR}]=1-F_l(\mathrm{VaR})$$

对于一个小 p，空头头寸的 VaR 为一个正值，正号表示一个损失.

前面的定义显示 VaR 关心的是 CDF $F_l(x)$的尾部行为. 对一个多头头寸，$F_l(x)$的左尾是重要的，然而对空头头寸则注意力集中于 $F_l(x)$的右尾. 注意到如果利用 $-\Delta V(l)$的分布，则方程(7.1)中 VaR 的定义也可以应用于空头头寸，因此，利用多头头寸来讨论 VaR 计算的方法就足够了.

对任何一元的 CDF $F_l(x)$与概率 p，满足 $0<p<1$，称

$$x_p=\inf\{x \mid F_l(x)\geqslant p\}$$

为 $F_l(x)$的 p 分位数，其中 inf 表示满足 $F_l(x)\geqslant p$ 的最小实数. 如果已知方程(7.1)中的 CDF $F_l(x)$，则 VaR 就是它的 p 分位数（即 $\mathrm{VaR}=x_p$）. 然而实际中 CDF 是未知的，从而 VaR 的研究关心的是 CDF 及其分位数的估计，尤其是 CDF 的尾部性质.

在实际应用中，VaR 的计算涉及几种因素：

1)感兴趣的概率 p. 例如 $p=0.01$ 或 $p=0.05$.

2)时间区间 l. 它可由管理委员会设定，如 1 天或 10 天.

3)数据频率．它可能与时间区间 l 是不一样的，经常使用的日观测值.

4)CDF $F_l(x)$或它的分位数.

5)金融头寸的数量或者组合的盯市价值 (mark-to-market value).

在这些因素中，CDF $F_l(x)$是经济计量模型的焦点，CDF 估计的不同方法引起了 VaR 计算的不同方法.

257~258

注释：方程(7.1)中的 VaR 定义是用美元度量的．因为对数收益率近似等于金融头寸价值的百分比变化，所以在数据分析中我们使用的是对数收益率 r_t．给定 t 时刻已知信息的条件下，由 r_{t+1}分布的分位数计算的 VaR 是用百分比表示的，因此，VaR 的美元计量为金融头寸的现金值乘以对数收益率序列的 VaR.

注释：VaR 关注的是组合在给定时间区间上的可能损失的一个预测，可以利用金融头寸将来收益率的预测分布来计算．例如，运用日收益率 r_t，组合在 1 天持有期的 VaR 可以利用 r_{t+1}在给定 t 时刻已知信息下的预测分布来计算．由统计的观点，在一个恰当指定的模型中，预测分布考虑的参数的不确定性．然而，预测分布是很难得到的，并且大多数 VaR 计算的可以利用的方法都忽视了参数不确定性的影响.

7.2 风险度量制

J. P. Morgan 将风险度量制方法(RiskMetrics™)发展到 VaR 计算中，参见 Longerstaey 和 More (1995)．风险度量制的简单形式假定组合的连续复合日收益率服从一个条件正态分布．用 r_t 表示日对数收益率，F_{t-1}表示 $t-1$ 时刻可以得到的信息集合，则风险度量制假定 $r_t \mid F_{t-1} \sim N(\mu_t, \sigma_t^2)$，其中 μ_t 是 r_t 的条件均值，σ_t^2是 r_t 的条件方差．另外，这个方法假定这两个量是随时间变化的，对应于简单的模型：

$$\mu_t = 0, \quad \sigma_t^2 = \alpha\sigma_{t-1}^2 + (1-\alpha)r_{t-1}^2, \qquad 0 < \alpha < 1 \tag{7.2}$$

因此，此方法假定组合的日价格的对数 $p_t = \ln(P_t)$，满足差分方程 $p_t - p_{t-1} = a_t$，其中 $a_t = \sigma_t \varepsilon_t$是一个无漂移的 IGARCH(1，1)过程．$\alpha$ 的值通常取在区间(0.9，1)上.

这样一个特殊的随机游走 IGARCH 模型的良好性质是：利用它很容易得到一个多期收益率的条件分布．具体地，对 k 个周期的持有期，从时刻 $t+1$ 到时刻 $t+k$(包含 $t+k$ 时刻)的对数收益率为

$$r_t[k] = r_{t+1} + \cdots + r_{t+k-1} + r_{t+k}$$

方括号[k]表示 k 期收益率．在方程(7.2)中具体的 IGARCH(1，1)模型下，条件分布 $r_t[k] \mid F_t$ 是均值为 0，方差为 $\sigma_t^2[k]$的正态分布，其中 $\sigma_t^2[k]$可以利用第 3 章中讨论的预测方法计算．由 ε_t 的独立性假定和模型(7.2)，我们有

$$\sigma_t^2[k] = \mathrm{Var}(r_t[k] \mid F_t) = \sum_{i=1}^{k} \mathrm{Var}(a_{t+i} \mid F_t)$$

其中 $\mathrm{Var}(a_{t+i} \mid F_t) = E(\sigma_{t+i}^2 \mid F_t)$可以递推地得到．利用 $r_{t-1} = a_{t-1} = \sigma_{t-1}\varepsilon_{t-1}$，我们可以将方程(7.2)中 IGARCH(1，1)模型的波动率方程改写为

259

$$\sigma_t^2=\sigma_{t-1}^2+(1-\alpha)\sigma_{t-1}^2(\epsilon_{t-1}^2-1) \quad 对所有\ t.$$

特别地，我们有

$$\sigma_{t+i}^2=\sigma_{t+i-1}^2+(1-\alpha)\sigma_{t+i-1}^2(\epsilon_{t+i-1}^2-1) \quad 对\ i=2,\cdots,k.$$

因为对 $i\geqslant 2$，$E(\epsilon_{t+i-1}^2-1 \mid F_t)=0$，所以前面的方程说明

$$E(\sigma_{t+i}^2 \mid F_t)=E(\sigma_{t+i-1}^2 \mid F_t) \quad 对\ i=2,\cdots,k. \tag{7.3}$$

对向前 1 步的波动率预测，由方程(7.2) 知 $\sigma_{t+1}^2=\alpha\sigma_t^2+(1-\alpha)r_t^2$. 因此，方程 (7.3) 证明了对 $i\geqslant 1$，$\mathrm{Var}(r_{t+i} \mid F_t)=\sigma_{t+1}^2$，从而 $\sigma_t^2(k)=k\sigma_{t+1}^2$. 结果说明 $r_t[k] \mid F_t\sim N(0,\ k\sigma_{t+1}^2)$，因此，在方程(7.2)这个特殊的 IGARCH(1，1)模型下，$r_t[k]$的条件方差与时间段 k 成比例. k 持有期对数收益率的条件标准差为$\sqrt{k}\sigma_{t+1}$.

假定金融头寸是一个多头头寸，当有一个大的价格下降（即一个大的负收益率）时损失发生. 如果设定概率为 5%，则风险度量制用 $1.65\sigma_{t+1}$来度量组合风险，即它利用均值为 0，标准差为 σ_{t+1}的正态分布的单侧 5%分位数. 真实 5%分位数是$-1.65\sigma_{t+1}$，但是忽略负号，并理解为它代表一个损失. 因此，如果标准差是用百分比度量的，则风险度量制下组合的日 VaR 为

$$\mathrm{VaR}=头寸数量\times 1.65\sigma_{t+1}$$

k 天持有期的 VaR 为

$$\mathrm{VaR}(k)=头寸数量\times 1.65\sqrt{k}\,\sigma_{t+1}$$

其中 VaR 的变量(k)用来表示时间段. 因此，在风险度量制下，我们有

$$\mathrm{VaR}(k)=\sqrt{k}\times \mathrm{VaR}$$

这称为风险度量制下 VaR 计算的时间平方根法则.

例 7.1 1997 年 6 月德国马克/美元的汇率的连续复合日收益率的样本标准差大约是 260 0.53%. 假定一个投资者长期持有价值 1 千万美元的马克/美元汇率合约，则该投资者 1 天持有期的 5%VaR 是

$$10\,000\,000\times(1.65\times 0.005\,3)=87\,450(美元)$$

相应的 1 个月持有期（30 天）的 VaR 为

$$10\,000\,000\times(\sqrt{30}\times 1.65\times 0.005\,3)\approx 478\,983(美元)$$

例 7.2 考虑图 7-1 中 IBM 的日对数收益率. 如第 1 章中提到的，收益率的样本均值显著不同于 0. 然而，为了说明利用风险度量制的 VaR 计算，我们假定这个样本的条件均值为 0，而且收益率的波动率服从一个无漂移的 IGARCH(1，1)模型. 拟合的模型为

$$r_t=a_t,\quad a_t=\sigma_t\epsilon_t,\quad \sigma_t^2=0.939\,6\sigma_{t-1}^2+(1-0.939\,6)a_{t-1}^2 \tag{7.4}$$

其中$\{\epsilon_t\}$是标准的高斯白噪声序列. 如所料想的，由 Q 统计量，拒绝了这个模型. 例如，对标准残差的平方，我们有一个高度显著的统计量 $Q(10)=56.19$.

由数据与拟合的模型，我们有 $r_{9\,190}=-0.012\,8$，$\hat{\sigma}_{9\,190}^2=0.000\,347\,2$. 因此，1 步向前波动率预测为$\hat{\sigma}_{9\,190}^2(1)=0.000\,336$. 条件分布 $r_{9\,191} \mid F_{9\,190}$的 5%分位数为 $-1.65\times\sqrt{0.000\,336}=-0.030\,25$，这里将负号理解为表示损失. 因此，1 千万美元的多头头寸在 1 天持有期内的 5%VaR 为

$$\text{VaR} = 10\,000\,000 \times 0.030\,25 = 302\,500(\text{美元})$$

1%分位数是$-2.326\,2\times\sqrt{0.000\,336}=-0.042\,65$，相应的同样多头头寸的1%VaR 为 426 500 美元.

7.2.1 讨论

风险度量制的一个优点就是简单，易于理解和运用；另外一个优点是使得金融市场中的风险更加透明. 然而，因为证券收益率常常有厚尾（或肥尾），所以正态性假定通常导致 VaR 的低估. 其他计算 VaR 的方法要避免作这样一个假定.

时间平方根法则是风险度量制中运用的特殊模型的一个结果. 如果对数收益率的零均值假定或者对数收益率具体的 IGARCH(1，1)模型假设不满足，则此法则就失效了. 考虑下面这个简单模型：

$$r_t = \mu + a_t,\quad a_t = \sigma_t \epsilon_t,\quad \mu \neq 0,\quad \sigma_t^2 = \alpha\sigma_{t-1}^2 + (1-\alpha)a_{t-1}^2,$$

261

其中$\{\epsilon_t\}$是标准的高斯白噪声序列. $\mu\neq 0$ 的假定对许多在 NYSE 中大量交易的股票收益率都是成立的，见第 1 章. 对上面这个简单模型，给定 F_t 下 r_{t+1}的分布是 $N(\mu,\ \sigma_{t+1}^2)$. 用来计算 1 个持有期 VaR 的 5%分位数变为 $\mu-1.65\sigma_{t+1}$. 对 k 持有期，给定 F_t 下 $r_t[k]$的分布为 $N(k\mu,\ k\sigma_{t+1}^2)$，同前面一样，其中 $r_t[k]=r_{t+1}+\cdots+r_{t+k-1}+r_{t+k}$. k 持有期 VaR 计算中运用的 5%分位数是 $k\mu-1.65\sqrt{k}\,\sigma_{t+1}=\sqrt{k}(\sqrt{k}\,\mu-1.65\sigma_{t+1})$. 因此，当平均收益率不为 0 时，$\text{VaR}(k)\neq\sqrt{k}\times\text{VaR}$. 也很容易证明，当收益率的波动率模型不是无漂移的 IGARCH(1，1) 模型时，这个方法也是失效的.

7.2.2 多个头寸

在一些应用中，投资者可能持有多个头寸，并且需要计算头寸的全部 VaR. 做这样一个计算时，在假定每个头寸的日对数收益率服从一个随机游走 IGARCH(1，1)模型，风险度量制采取了一个简单方法. 需要的额外量是收益率间的交叉相关系数. 考虑两个头寸的情况. 令 VaR_1 和 VaR_2 表示两个头寸的 VaR，ρ_{12}表示两个收益率间的交叉相关系数，即

$$\rho_{12} = \text{Cov}(r_{1t}, r_{2t})/[\text{Var}(r_{1t})\text{Var}(r_{2t})]^{0.5}$$

则投资者的全部 VaR 为

$$\text{VaR} = \sqrt{\text{VaR}_1^2 + \text{VaR}_2^2 + 2\rho_{12}\text{VaR}_1\text{VaR}_2}$$

一个包含 m 个工具的头寸的 VaR 可以直接得到，为

$$\text{VaR} = \sqrt{\sum_{i=1}^{m}\text{VaR}_i^2 + 2\sum_{i<j}^{m}\rho_{ij}\text{VaR}_i\text{VaR}_j},$$

其中 ρ_{ij}是第 i 个与第 j 个工具的收益率间的交叉相关系数，VaR_i 表示第 i 个工具的 VaR.

7.3 VaR 计算的经济计量方法

VaR 计算的一个一般方法是利用第 2 章到第 4 章中时间序列的经济计量模型. 对一个对数收益率序列，可以利用第 2 章中的时间序列模型对均值方程建模，利用第 3 章或第 4 章的条件

异方差模型处理波动率. 为了简便，我们在讨论中利用 GARCH 模型，并将此方法称为 VaR 计算的经济计量方法. 也可以利用其他的波动率模型，包括第4章中的非线性模型.

262

考虑资产的对数收益率 r_t，它的一般时间序列模型可以写为

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j} \tag{7.5}$$

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{u} \alpha_i a_{t-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t-j}^2 \tag{7.6}$$

方程(7.5)和方程(7.6)是 r_t 的均值和波动率方程. 假定参数已知时，可以利用这两个方程来得到 r_t 的条件均值与条件方差的1步向前预测. 具体地，我们有

$$\hat{r}_t(1) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t+1-i} - \sum_{j=1}^{q} \theta_j a_{t+1-j}$$

$$\hat{\sigma}_t^2(1) = \alpha_0 + \sum_{i=1}^{u} \alpha_i a_{t+1-i}^2 + \sum_{j=1}^{v} \beta_j \sigma_{t+1-j}^2$$

如果进一步假定 ε_t 是高斯的，则给定 t 时刻可得信息的条件下，r_{t+1} 的条件分布是 $N[\hat{r}_t(1), \hat{\sigma}_t^2(1)]$. 用 VaR 计算的这个条件分布的分位数可以很容易得到. 例如，其5%分位数是 $\hat{r}_t(1) - 1.65\,\hat{\sigma}_t(1)$. 如果假定 ε_t 是自由度为 v 的标准化的学生-t 分布，则分位数为 $\hat{r}_t(1) - t_v^*(p)\hat{\sigma}_t(1)$，其中 $t_v^*(p)$ 表示自由度为 v 的标准化学生-t 分布的 p 分位数.

自由度为 v 的学生-t 分布(用 t_v 表示)的分位数与它的标准化分布(用 t_v^* 表示)的分位数之间有如下关系:

$$p = \Pr(t_v \leqslant q) = \Pr\left(\frac{t_v}{\sqrt{v/(v-2)}} \leqslant \frac{q}{\sqrt{v/(v-2)}}\right) = \Pr\left(t_v^* \leqslant \frac{q}{\sqrt{v/(v-2)}}\right)$$

其中 $v>2$. 也就是说，如果 q 是自由度为 v 的学生-t 分布的 p 分位数，则 $q/\sqrt{v/(v-2)}$ 就是自由度为 v 的标准化学生-t 分布的 p 分位数. 因此，如果方程(7.6)的 GARCH 模型中的 ε_t 服从自由度为 v 的标准化学生-t 分布，概率是 p，则在时刻 t 时，用来计算1持有期 VaR 的分位数为

$$\hat{r}_t(1) - \frac{t_v(p)\,\hat{\sigma}_t(1)}{\sqrt{v/(v-2)}}$$

263

其中 $t_v(p)$ 是自由度为 v 的学生-t 分布的 p 分位数.

例 7.3 再次考虑例7.2中 IBM 的日对数收益率. 我们利用两个波动率模型来计算一个多头头寸1千万美元在 $t=9\,190$ 时1天持有期的 VaR. 根据第2章与第3章中的建模方法，这些经济计量模型是合理的.

情形1

假定 ε_t 是标准正态的，则拟合的模型为

$$r_t = 0.000\,66 - 0.024\,7 r_{t-2} + a_t, \quad a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = 0.000\,003\,89 + 0.079\,9 a_{t-1}^2 + 0.907\,3 \sigma_t^2$$

由数据，我们有 $r_{9\,189}=-0.002\ 01$，$r_{9\,190}=-0.012\ 8$，$\sigma^2_{9\,190}=0.000\ 334\ 55$. 因此，前面的 AR(2)-GARCH(1，1)模型产生的 1 步向前预测为

$$\hat{r}_{9\,190}(1)=0.000\ 71, \hat{\sigma}^2_{9\,190}(1)=0.000\ 321\ 1$$

5％分位数为

$$0.000\ 71-1.644\ 9\times\sqrt{0.000\ 321\ 1}=-0.028\ 77$$

这里负号理解为表示条件正态分布的左尾. 多头头寸 1 千万美元概率为 0.05 的 VaR 为 VaR＝10 000 000×0.028 77＝287 700 美元. 结果显示，假定 AR(2)-GARCH(1，1)模型成立，则以概率 95％，下一天持有此头寸的潜在损失是 287 200 美元或低于这个值. 如果概率为 0.01，那么 1％分位数为

$$0.000\ 71-2.326\ 2\times\sqrt{0.000\ 321\ 1}=-0.040\ 973\ 8$$

头寸的 VaR 变为 409 738 美元.

情形 2

假定 ϵ_t 是自由度为 5 的标准化学生-t 分布，拟合的模型为

$$r_t=0.000\ 3-0.033\ 5r_{t-2}+a_t,\quad a_t=\sigma_t\epsilon_t$$
$$\sigma_t^2=0.000\ 003+0.055\ 9a_{t-1}^2+0.935\ 0\sigma_{t-1}^2$$

由数据，我们有 $r_{9\,189}=-0.002\ 01$，$r_{9\,190}=-0.012\ 8$，$\sigma^2_{9\,190}=0.000\ 349$. 因此，前面的学生-$t$ AR(2)-GARCH(1，1)模型产生的 1 步向前预测为

$$\hat{r}_{9\,190}(1)=0.000\ 367,\qquad \hat{\sigma}^2_{9\,190}(1)=0.000\ 338\ 6.$$ [264]

自由度为 5 的学生-t 分布的 5％分位数为－2.015，它的标准化分布的 5％分位数为$-2.015/\sqrt{5/3}=-1.560\ 8$. 因此，给定 $F_{9\,190}$条件下，$r_{9\,191}$的条件分布的 5％分位数为

$$0.000\ 367-1.560\ 8\sqrt{0.000\ 338\ 6}=-0.028\ 354$$

一个 1 千万美元的多头头寸的 VaR 为

$$\text{VaR}=10\ 000\ 000\times 0.028\ 352=283\ 520(\text{美元})$$

这基本上与正态性假定下得到的结果相同. 条件分布的 1％分位数为

$$0.000\ 367-(3.364\ 9/\sqrt{5/3})\sqrt{0.000\ 338\ 6}=-0.047\ 594\ 3$$

对应的 VaR 为 475 943 美元. 与情形 1 的结果比较，我们可以看出利用自由度为 5 的学生-t 分布时的厚尾效应；当尾概率变小时增加了 VaR.

多个周期

假定 h 时刻，我们希望计算对数收益率为 r_t 的资产的 k 期 VaR. 感兴趣的变量是预测原点 h 处的 k 期对数收益率(即 $r_h[k]=r_{h+1}+\cdots+r_{h+k}$). 如果收益率 r_t 服从方程(7.5)和方程(7.6)中的时间序列模型，则可以通过第 2 章和第 3 章中讨论的预测方法来得到 $r_h[k]$在给定信息集 F_h 下的条件均值与条件方差.

期望收益率与预测误差

可以利用第 2 章中 ARMA 模型的预测方法来得到条件均值 $E(r_h[k]\mid F_h)$. 具体地，我们有

$$\hat{r}_h[k] = r_h(1) + \cdots + r_h(k)$$

其中 $r_h(l)$是收益率在预测原点 h 的 l 步向前预测，这些预测可以利用 2.6.4 节讨论的递推方法来计算．利用方程(7.5)中 ARMA 模型的 MA 表示

$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots,$$

我们可以将预测原点 h 处的 l 步向前预测误差写为

265

$$e_h(l) = r_{h+l} - r_h(l) = a_{h+l} + \psi_1 a_{h+l-1} + \cdots + \psi_{l-1} a_{h+1}.$$

参见方程(2.30)以及相应的预测误差．k 期期望收益率的预测误差$\hat{r}_h[k]$是 r_t 在预测原点 h 处的 1 步向前预测误差的和，可以写为

$$\begin{aligned} e_h[k] &= e_h(1) + e_h(2) + \cdots + e_h(k) \\ &= a_{h+1} + (a_{h+2} + \psi_1 a_{h+1}) + \cdots + \sum_{i=0}^{k-1} \psi_i a_{h+k-i} \\ &= a_{h+k} + (1+\psi_1) a_{h+k-1} + \cdots + \left(\sum_{i=0}^{k-1} \psi_i\right) a_{h+1} \end{aligned} \tag{7.7}$$

其中 $\psi_0 = 1$.

期望波动率

k 期收益率在预测原点 h 处的波动率预测是 $e_h[k]$ 在给定 F_h 下的条件方差．利用对 $i=1, \cdots, k$，ε_{t+i}的独立性假定，且 $a_{t+i} = \sigma_{t+i}\varepsilon_{t+i}$，我们有

$$\begin{aligned} \mathrm{Var}(e_h[k] \mid F_h) &= \mathrm{Var}(a_{h+k} \mid F_h) + (1+\psi_1)^2 \mathrm{Var}(a_{h+k-1} \mid F_h) + \cdots \\ &\quad + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \mathrm{Var}(a_{h+1} \mid F_h) \\ &= \sigma_h^2(k) + (1+\psi_1)^2 \sigma_h^2(k-1) + \cdots + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \sigma_h^2(1) \end{aligned}$$

其中 $\sigma_h^2(l)$是预测原点 h 处的 l 步向前波动率预测．如果波动率模型是方程(7.6)中的 GARCH 模型，那么这些波动率预测可以由第 3 章讨论的方法递推得到．

作为说明，考虑特殊的时间序列模型

$$r_t = \mu + a_t, \quad a_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

则对所有的 $i>0$，有 $\psi_i = 0$．k 期收益率在预测原点 h 处的点预测为$\hat{r}_h[k] = k\mu$，对应的预测误差为

$$e_h[k] = a_{h+k} + a_{h+k-1} + \cdots + a_{h+1}.$$

因此，k 期收益率在预测原点 h 处的波动率预测为

266

$$\mathrm{Var}(e_h[k] \mid F_h) = \sum_{l=1}^{k} \sigma_h^2(l).$$

利用第 3.4 节中 GARCH(1，1)模型的预测方法，我们有

$$\begin{aligned} \sigma_h^2(1) &= \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2 \\ \sigma_h^2(l) &= \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(l-1), \qquad l = 2, \cdots, k. \end{aligned} \tag{7.8}$$

这样，$\mathrm{Var}(e_h[k] \mid F_h)$可以通过前面的递推得到．如果 ϵ_t 是高斯的，则 $r_h[k]$在给定 F_h 下的条件分布是均值为 $k\mu$，方差为 $\mathrm{Var}(e_h[k] \mid F_h)$的正态分布．VaR 计算中需要的分位数很容易得到．

例 7.3（续） 对于 IBM 股票的日对数收益率，考虑例 7.3 中的高斯 AR(2)-GARCH(1, 1)模型．假定我们感兴趣的是在预测原点 9190（即 1998 年 12 月 31 日）开始的 15 天持有期的 VaR．我们可以利用拟合的模型来计算 15 天对数收益率的条件均值和方差，通过给定 $F_{9\,190}$，$r_{9\,190}[15]=\sum_{i=1}^{15} r_{9\,190+i}$．由方程(7.8)递推得到条件均值为 0.009 98，条件方差为 0.004 794 8．那么条件分布的 5%分位数为 $0.009\,98-1.644\,9\sqrt{0.004\,794\,8}=-0.103\,919\,1$．因此，一个 1 千万美元的多头头寸的 15 天持有期的 VaR 为

$$\mathrm{VaR} = 10\,000\,000 \times 0.103\,919\,1 = 1\,039\,191(\text{美元})$$

这个数低于 $287\,700\times\sqrt{15}=1\,114\,257$ 美元．由这个例子进一步表明，风险度量制使用的平方根时间法则仅仅对运用特殊的白噪声 IGARCH(1，1)模型成立．当条件均值不为 0 时，必须采取恰当的步骤计算 k 持有期的 VaR.

7.4 分位数估计

分位数估计提供了 VaR 计算的非参数方法．除了假定该分布在预测阶段仍然成立以外，它不对组合的收益率作具体的分布假定．有两种类型的分位数方法：第一种方法是直接利用经验分位数；第二种方法是运用分位数回归.

7.4.1 分位数与次序统计量

假定收益率的分布在预测期间与样本期间是一样的，可以利用收益率 r_t 的经验分位数来计算 VaR．令 r_1，…，r_n 表示样本期间内组合的收益率．样本的次序统计量是这些值用递增次序排列后的值．我们利用记号

$$r_{(1)} \leqslant r_{(2)} \leqslant \cdots \leqslant r_{(n)}$$

表示这个排列，并将 $r_{(i)}$称为样本的第 i 个次序统计量．特别地，$r_{(1)}$表示样本极小值，$r_{(n)}$表示样本极大值.

267

假定收益率是独立同分布并有着一个连续分布的随机变量，其概率密度函数(pdf) $f(x)$和 CDF $F(x)$，那么由统计文献（例如 Cox 和 Hinkley，1974，见附录 2），对次序统计量 $r_{(l)}$，其中 $l=np$，$0<p<1$，我们有下面的渐近结果.

结果： *令 x_p 表示 $F(x)$的 p 分位数[即 $x_p=F^{-1}(p)$]．假定 pdf $f(x)$ 在 x_p 处不等于 0(即 $f(x_p)\neq 0$)，则次序统计量 $r_{(l)}$ 是渐近正态的，均值为 x_p，方差为 $p(1-p)/[nf^2(x_p)]$．也就是，*

$$r_{(l)} \sim N\left[x_p, \frac{p(1-p)}{n[f(x_p)]^2}\right], \qquad l = np. \tag{7.9}$$

根据前面的结果，可以利用 $r_{(l)}$ 估计分位数 x_p，这里 $l=np$. 实际中，感兴趣的概率 p 可能不满足 np 是正整数的条件. 在这种情况下，可以利用简单的插值得到分位数估计. 更具体地，对非整数 np，令 l_1 和 l_2 表示与 np 最邻近的两个正整数，满足 $l_1<np<l_2$. 并定义 $p_i=l_i/n$. 前面的结果证明了 $r_{(l_i)}$ 是分位数 x_{p_i} 的一个相合估计. 由定义，$p_1<p<p_2$，因此，分位数 x_p 可以通过下式估计

$$\hat{x}_p=\frac{p_2-p}{p_2-p_1}r_{(l_1)}+\frac{p-p_1}{p_2-p_1}r_{(l_2)}. \tag{7.10}$$

例 7.4 考虑英特尔股票从 1972 年 12 月 15 日至 1997 年 12 月 31 日的日对数收益率，共有 6 329 个观测值. 可以得到数据 5%的经验分位数为

$$\hat{x}_{0.05}=0.55r_{(316)}+0.45r_{(317)}=-4.229\%,$$

这里 $np=6\,329\times0.05=316.45$，$r_{(i)}$ 表示样本的第 i 个次序统计量. 在这个特例中，$r_{(316)}=-4.237\%$，$r_{(317)}=-4.220\%$. 这里我们利用了经验分布的左尾 (lower tail)，因为它在 VaR 的计算中与持有一个多头头寸有关.

例 7.5 再考虑 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日的日对数收益率. 利用全部的 9 190 个观测值，可以得到 5%的经验分位数为 $(r_{(459)}+r_{(460)})/2=-0.021\,603$，其中 $r_{(i)}$ 为第 i 个次序统计量，而且 $np=9\,190\times0.05=459.5$. 1 千万美元的多头头寸的 VaR 为 216 030 美元，这远远小于前面用经济计量方法得到的结果. 因为样本大小为 9 190，我们有 $91<9\,190\times0.01<92$，令 $p_1=91/9\,190=0.009\,9$，$p_2=92/9\,190=0.010\,01$，则得到 1%的经验分位数为

268

$$\begin{aligned}\hat{x}_{0.01}&=\frac{p_2-0.01}{p_2-p_1}r_{(91)}+\frac{0.01-p_1}{p_2-p_1}r_{(92)}\\&=\frac{0.000\,01}{0.000\,11}(-3.658)+\frac{0.000\,1}{0.000\,11}(-3.657)\\&\approx-3.657\end{aligned}$$

这个多头头寸的 1%的 1 天持有期的 VaR 为 365 709 美元，又一次低于前面其他方法得到的结果.

讨论： 利用上面的 VaR 计算的分位数方法有几个优势，包括(a)简单性，(b)没有具体的分布假定. 然而，这个方法也有几个缺点. 首先，它假定收益率 r_t 的分布从样本期间到预测期间是保持不变的. 假设 VaR 主要关心的是尾概率，这个假定蕴含了预测的损失不能高于历史的损失，然而实际中并不是这样的. 第二，对极端分位数(即当 p 接近于 0 或 1 时)，经验分位数并非是理论分位数的有效估计. 第三，直接的分位数估计不能考虑与所研究的组合相关的解释变量的影响. 在实际应用中，由经验分位数得到的 VaR 可以充当真实 VaR 的一个下界.

7.4.2 分位数回归

实际应用中，人们经常能够得到对所研究的问题非常重要的解释变量. 例如，FRB(美国联邦储备银行)对利率采取的行动对美国股票的收益率有重要的影响，从而考虑分布函数 $r_{t+1}\mid F_t$ (这里 F_t 包含了这个解释变量)是更加恰当的. 换句话说，我们对给定 F_t 下 r_{t+1} 的分布函数的

分位数感兴趣. 这样一个分位数在文献中称为回归分位数，可以参见 Koenker 和 Bassett (1978).

为了理解回归分位数，将前一小节中的经验分位数看作一个估计问题是有帮助的. 对一个给定的概率 p，$\{r_t\}$的 p 分位数可以通过下式得到

$$\hat{x}_p = \arg\min_{\beta} \sum_{i=1}^{n} w_p(r_i - \beta)$$

其中 $w_p(z)$定义为

$$w_p(z) = \begin{cases} pz & 若\ z \geqslant 0 \\ (p-1)z & 若\ z < 0 \end{cases},$$

回归分位数是这样一个估计的一般化.

269

为了看出这种一般化，假定我们有线性回归

$$r_t = \boldsymbol{\beta}'\boldsymbol{x}_t + a_t \tag{7.11}$$

其中$\boldsymbol{\beta}$是k 维参数向量；$\boldsymbol{x}_t$ 是预测向量，它是 F_{t-1}的元素. 因为$\boldsymbol{\beta}'\boldsymbol{x}_t$ 是已知的，所以 r_t 在给定 F_{t-1}条件下的条件分布就是 a_t 分布的一个平移. 用这种方式看待这个问题，Koenker 和 Bassett (1978) 指出 r_t 在给定 F_{t-1}条件下的条件分位数 $x_p \mid F_{t-1}$的估计为

$$\hat{x}_p \mid F_{t-1} \equiv \inf\{\boldsymbol{\beta}_0'\boldsymbol{x} \mid R_p(\boldsymbol{\beta}_0) = \min\} \tag{7.12}$$

其中“$R_p(\boldsymbol{\beta}_0)=\min$”意思是$\boldsymbol{\beta}_0$ 可以由下式得到

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{n} w_p(r_t - \boldsymbol{\beta}'\boldsymbol{x}_t),$$

其中 $w_p(.)$如前面定义. 得到这样一个估计分位数的计算机程序可以在 Koenker 和 D'Orey (1987) 中找到.

7.5 极值理论

本节中，我们回顾统计文献中的一些极值理论. 用 r_t 表示一个资产以一个给定时间间隔(如天)测量的收益率. 考虑 n 个收益率的集合$\{r_1, \cdots, r_n\}$，该集合中的最小收益率为 $r_{(1)}$，也就是最小次序统计量；而最大收益率 $r_{(n)}$，也就是最大次序统计量. 具体地，$r_{(1)} = \min_{1\leqslant j\leqslant n}\{r_j\}$，$r_{(n)} = \max_{1\leqslant j\leqslant n}\{r_j\}$. 对于一个多头头寸，最小收益率 $r_{(1)}$与 VaR 的计算是高度相关的，所以我们主要讨论 $r_{(1)}$的性质. 然而，所讨论的理论对一个资产在给定时间间隔上的最大收益率也是适用的，因为最大收益率的性质可以通过最小收益率的一个简单的符号变化得到. 具体地，我们有 $r_{(n)} = -\min_{1\leqslant j\leqslant n}\{-r_j\} = -r^c_{(1)}$，其中 $r^c_t = -r_t$，上标 c 表示符号的变化. 最大收益率与持有一个空头金融头寸有关.

7.5.1 极值理论的回顾

假设收益率 r_t 是前后独立的，其通常的 CDF 为 $F(x)$，收益率 r_t 的变化范围为$[l, u]$. 对于对数收益率，我们有 $l=-\infty$，$u=\infty$. $r_{(1)}$的 CDF(用 $F_{n,1}(x)$表示)由下式给出

270

$$
\begin{aligned}
F_{n,1}(x) &= \Pr[r_{(1)} \leqslant x] = 1 - \Pr[r_{(1)} > x] \\
&= 1 - \Pr(r_1 > x, r_2 > x, \cdots, r_n > x) \\
&= 1 - \prod_{j=1}^{n} \Pr(r_j > x) \quad (\text{由独立性}) \\
&= 1 - \prod_{j=1}^{n} [1 - \Pr(r_j \leqslant x)] \\
&= 1 - \prod_{j=1}^{n} [1 - F(x)] \quad (\text{由通常的分布}) \\
&= 1 - [1 - F(x)]^n
\end{aligned}
\tag{7.13}
$$

实际中，r_t 的 CDF $F(x)$是未知的. 因此，$r_{(1)}$的 $F_{n,1}(x)$也是未知的. 然而，当 n 趋于无穷时，$F_{n,1}(x)$变成退化的. 即当 n 趋于无穷时，若 $x \leqslant l$，则 $F_{n,1}(x) \to 0$；若 $x > l$，则 $F_{n,1}(x) \to 1$. 因为这个退化的 CDF 没有实际价值，所以极值理论关心的是寻找两个序列$\{\beta_n\}$和$\{\alpha_n\}$(其中 $\alpha_n > 0$)，满足 $r_{(1^*)} \equiv (r_{(1)} - \beta_n)/\alpha_n$ 的分布当 n 趋于无穷时收敛到一个非退化分布. 序列$\{\beta_n\}$是一个位置序列，$\{\alpha_n\}$是尺度因子序列. 在独立性假定下，标准化的最小收益率 $r_{(1^*)}$ 的极限分布为

$$
F_*(x) = \begin{cases} 1 - \exp[-(1 + kx)^{1/k}] & \text{若 } k \neq 0 \\ 1 - \exp[-\exp(x)] & \text{若 } k = 0 \end{cases}
\tag{7.14}
$$

若 $k<0$，则上式对 $x<-1/k$ 成立；若 $k>0$，则上式对 $x>-1/k$ 成立，其中下标 $*$ 表示最小收益率. $k=0$ 的情况看作为 $k \to 0$ 时的极限. 参数 k 称为形状参数，它控制了极限分布的尾行为. 参数 $\alpha = -1/k$ 称为分布的尾指数.

方程(7.14)的极限分布是 Jenkinson (1955) 对最小收益率的一般的极值分布. 它包含了 Gnedenko (1943)中三种类型的极限分布：

- 类型Ⅰ：$k=0$，冈贝尔(Gumbel)族，其 CDF 为

$$
F_*(x) = 1 - \exp[-\exp(x)], \qquad -\infty < x < \infty
\tag{7.15}
$$

- 类型Ⅱ：$k<0$，弗雷歇(Fréchet)族，其 CDF 为

$$
F_*(x) = \begin{cases} 1 - \exp[-(1 + kx)^{1/k}] & \text{若 } x < -1/k \\ 1 & \text{其他} \end{cases}
\tag{7.16}
$$

- 类型Ⅲ：$k>0$，韦布尔族，这里其 CDF 为

$$
F_*(x) = \begin{cases} 1 - \exp[-(1 + kx)^{1/k}] & \text{若 } x > -1/k \\ 0 & \text{其他} \end{cases}
$$

271

Gnedenko(1943)给出了与上面的三种类型极限分布之一相联系的 r_t 的 CDF $F(x)$的充要条件. 简要地说，$F(x)$的尾行为决定了最小收益率的极限分布 $F_*(x)$. 与冈贝尔族相联系的分布的(左)尾是指数衰减的；与弗雷歇族相联系的分布的(左)尾是一个幂函数；与韦布尔族相联系的分布的下端点是有限的. 对于极值理论的全面讨论，读者可以参阅 Embrechts，Kuppelberg 和 Mikosch (1997). 在风险管理中，我们主要对弗雷歇族感兴趣，它包含了平稳分布和学生-t 分

布. 冈贝尔族包含了如正态分布、对数正态分布等薄尾分布. 方程(7.14)的一般极限分布的概率分布密度函数(pdf)可以很容易地通过差分得到：

$$f_*(x)=\begin{cases}(1+kx)^{1/k-1}\exp[-(1+kx)^{1/k}] & \text{若 } k\neq 0\\ \exp[x-\exp(x)] & \text{若 } k=0\end{cases}\tag{7.17}$$

这里对 $k=0$，$-\infty<x<\infty$；对 $k<0$，$x<-1/k$；对 $k>0$，$x>-1/k$.

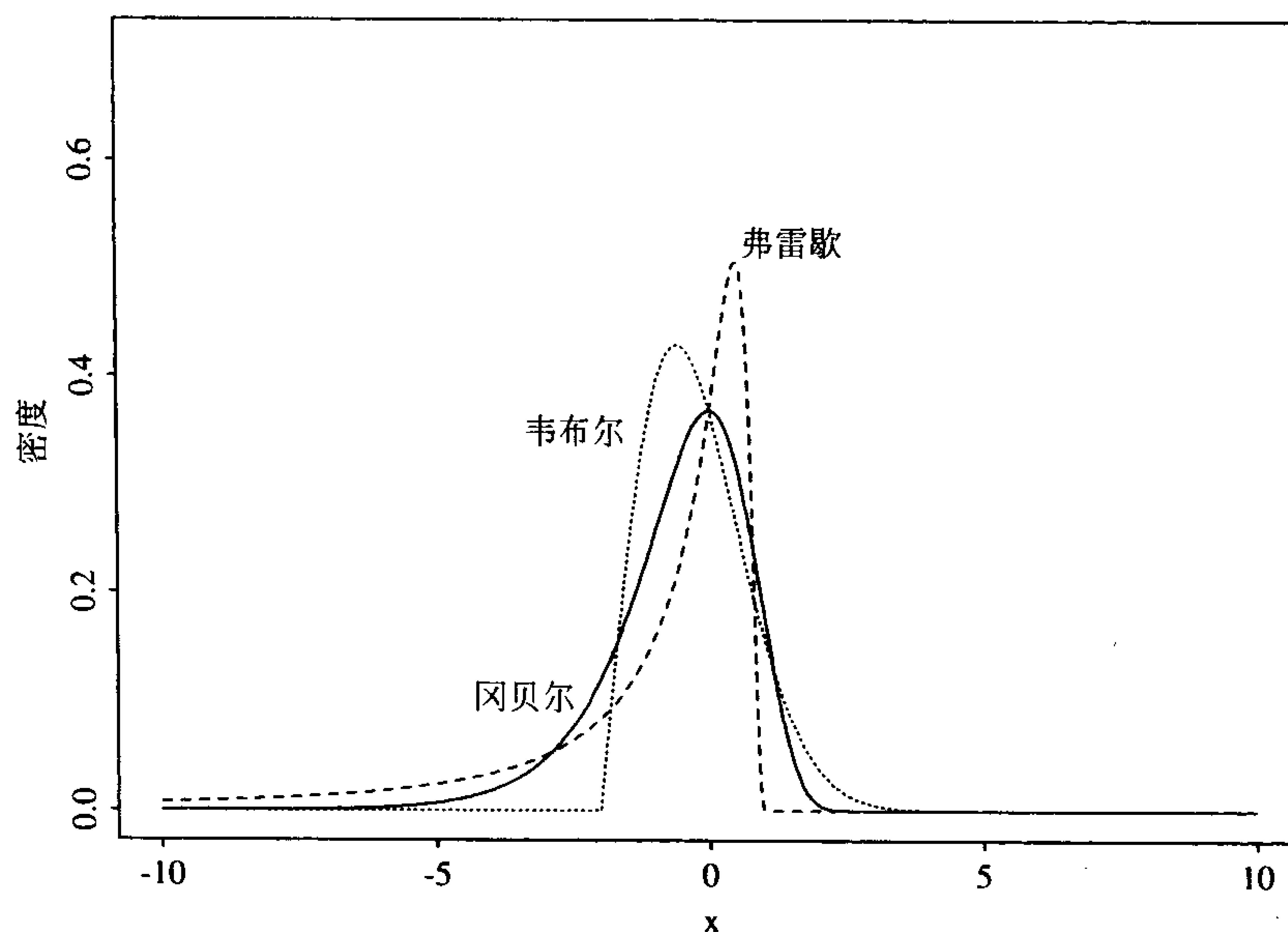


图 7-2　最小收益率的极值分布的概率密度函数：实线是冈贝尔分布，点虚线是 $k=0.5$ 的韦布尔分布，虚线是 $k=-0.9$ 的弗雷歇分布

前面提到的极值理论有两个重要的应用. 首先，r_t 的 CDF $F(x)$的尾部行为(而不是具体的分布)决定了(标准化的)最小收益率的极限分布 $F_*(x)$. 这样，此理论对于收益率 r_t 的一个广泛的分布范围都是实际可行的. 然而序列 $\{\beta_n\}$和$\{\alpha_n\}$可能依赖于 CDF $F(x)$. 第二，Feller (1971，第 279 页)证明了尾指数 k 并不依赖于 r_t 的时间间隔. 也就是说，尾指数(或等价地称为形状参数)在时间累积下是不变的. 极限分布的第二个性质在 VaR 计算中变得可以利用. 272

极值理论已经扩展到序列相关的观测值 $\{r_t\}_{t=1}^n$，假定这种相关是很弱的. Berman(1964)证明了假定 r_t 的自相关函数是平方可积的(即 $\sum_{i=1}^{\infty}\rho_i^2<\infty$)，其中 ρ_i 是 r_t 的延迟 i 的自相关函数，则极限极值分布的同样形式对平稳的正态序列也成立. 关于序列相关性对极值理论的影响的进一步结果，读者可以参考 Leadbetter，Lindgren 和 Rootzén (1983，第 3 章).

7.5.2　经验估计

极值分布包含 3 个参数：k，β_n，α_n. 这些参数分别称为形状参数、位置参数和尺度参数. 可以利用参数或非参数方法来估计，我们回顾一些估计方法.

对给定的样本，只有一个最小收益率或最大收益率，我们仅用一个极端观测不能估计三个参数，必须利用另外的方法. 文献中运用的一个想法是将样本分成子样，并对子样应用极值理论. 假定共有 T 个收益率 $\{r_j\}_{j=1}^T$，我们将样本分成 g 个互不相交的子样，每个子样有 n 个观测值，为了简便，假定 $T=ng$. 换句话说，我们将数据分为

$$\{r_1,\cdots,r_n \mid r_{n+1},\cdots,r_{2n} \mid r_{2n+1},\cdots,r_{3n} \mid \cdots \mid r_{(g-1)n+1},\cdots,r_{ng}\},$$

并将观测到的收益率记为 r_{in+j}，其中 $1\leqslant j\leqslant n$，$i=0$，…，$g-1$. 注意到每个子样对应于数据区间的一个子区间. 当 n 充分大时，我们希望极值理论对每个子样都适用. 应用时，n 的选择由实际情况来决定. 例如，对于日收益率，$n=21$ 近似对应于一个月内的交易日数量；$n=63$ 表示一个季度交易日的数量.

令 $r_{n,i}$ 表示第 i 个子样的最小值(即，$r_{n,i}$ 指第 i 个子样中的最小收益率)，其中下标 n 用来表示子样的大小. 当 n 充分大时，$x_{n,i}=(r_{n,i}-\beta_n)/\alpha_n$ 应该服从一个极值分布，并且子样最小值的集合 $\{r_{n,i} \mid i=1, \cdots, g\}$ 可以认为是从极值分布中抽取的 g 个观测的样本. 具体地，我们定义

$$r_{n,i}=\min_{1\leqslant j\leqslant n}\{r_{(i-1)n+j}\}, \qquad i=1,\cdots,g \tag{7.18}$$

子样最小值的集合 $\{r_{n,i}\}$ 是我们用来估计极值分布未知参数的数据. 显然，得到的估计可能依赖于子区间长度 n 的选择.

273

7.5.2.1 参数方法

有两种参数方法可以用，它们是最大似然法和回归方法.

最大似然法

假定子区间最小值 $\{r_{n,i}\}$ 服从一般的极值分布，满足 $x_i=(r_{n,i}-\beta_n)/\alpha_n$ 的 pdf 由方程(7.17)给出，我们可以通过简单的转移变换得到 $r_{n,i}$ 的 pdf 为

$$f(r_{n,i})=\begin{cases}\dfrac{1}{\alpha_n}\left[1+\dfrac{k_n(r_{n,i}-\beta_n)}{\alpha_n}\right]^{1/k_n-1}\exp\left\{-\left[1+\dfrac{k_n(r_{n,i}-\beta_n)}{\alpha_n}\right]^{1/k_n}\right\}, & \text{若 } k_n\neq 0\\ \dfrac{1}{\alpha_n}\exp\left\{\dfrac{r_{n,i}-\beta_n}{\alpha_n}-\exp\left[\dfrac{r_{n,i}-\beta_n}{\alpha_n}\right]\right\}, & \text{若 } k_n=0\end{cases}$$

这里理解为若 $k_n\neq 0$，则 $1+k_n(r_{n,i}-\beta_n)/\alpha_n>0$. 形状参数 k 中加入下标 n，表示它的估计依赖于 n 的选择. 在独立性假定下，子区间最小值的似然函数为

$$l(r_{n,1},\cdots,r_{n,g} \mid k_n,\alpha_n,\beta_n)=\prod_{i=1}^{g} f(r_{n,i})$$

可以利用非线性估计程序来得到 k_n，β_n，和 α_n 的最大似然估计. 这些估计是无偏的，渐近正态的，且在适当的假设下具有最小方差. 稍后我们对一些股票收益率序列应用这些方法.

回归方法

这个方法假定 $\{r_{n,i}\}_{i=1}^g$ 是从方程(7.14)的一般极值分布中抽取的一个随机样本，利用次序统计量的性质；见 Gumbel(1958). 将子区间最小值 $\{r_{n,i}\}_{i=1}^g$ 的次序统计量表示为

$$r_{n(1)}\leqslant r_{n(2)}\leqslant\cdots\leqslant r_{n(g)}.$$

利用次序统计量的性质(如 Cox and Hinkley，1974，第 467 页)，我们有

$$E\{F_*[r_{n(i)}]\}=\frac{i}{g+1},\qquad i=1,\cdots,g. \tag{7.19}$$

为了简便，我们根据 k 的值分为两种情形讨论．首先，考虑 $k\neq0$ 的情形．由方程(7.14)，我们有

274

$$F_*[r_{n(i)}]=1-\exp\left\{-\left[1+k_n\frac{r_{n(i)}-\beta_n}{\alpha_n}\right]^{1/k_n}\right\} \tag{7.20}$$

因此，利用方程(7.19)、(7.20)以及观测值的渐近期望，我们有

$$\frac{i}{g+1}=1-\exp\left\{-\left[1+k_n\frac{r_{n(i)}-\beta_n}{\alpha_n}\right]^{1/k_n}\right\}.$$

因此，

$$\exp\left\{-\left[1+k_n\frac{r_{n(i)}-\beta_n}{\alpha_n}\right]^{1/k_n}\right\}=1-\frac{i}{g+1}=\frac{g+1-i}{g+1},\qquad i=1,\cdots,g$$

连续两次取自然对数，前面的方程给出

$$\ln\left[-\ln\left(\frac{g+1-i}{g+1}\right)\right]=\frac{1}{k_n}\ln\left[1+k_n\frac{r_{n(i)}-\beta_n}{\alpha_n}\right],\qquad i=1,\cdots,g$$

实际中，令 e_i 表示前面两个量之间的偏移，并假设序列 $\{e_t\}$ 不是序列相关的，则我们有一个回归步骤

$$\ln\left[-\ln\left(\frac{g+1-i}{g+1}\right)\right]=\frac{1}{k_n}\ln\left[1+k_n\frac{r_{n(i)}-\beta_n}{\alpha_n}\right]+e_i,\qquad i=1,\cdots,g \tag{7.21}$$

可以通过最小化 e_i 的平方和来得到 k_n、β_n 与 α_n 的最小二乘估计．

当 $k_n=0$ 时，回归的建立简化为

$$\ln\left[-\ln\left(\frac{g+1-i}{g+1}\right)\right]=\frac{1}{\alpha_n}r_{n(i)}-\frac{\beta_n}{\alpha_n}+e_i,\qquad i=1,\cdots,g$$

最小二乘估计是相合的，但比最大似然估计的有效性低．本章中我们将利用最大似然估计．

7.5.2.2　非参数方法

形状参数 k 可以利用一些非参数方法来估计，这里我们将提到两种这样的方法．这两种方法是由 Hill(1975)和 Pickands(1975)提出的，称为 Hill 估计和 Pickands 估计．将两种估计都直接应用到收益率 $\{r_t\}_{t=1}^{T}$ 上，这样就没有必要考虑子样了．将样本的次序统计量表示为 $r_{(1)}\leqslant r_{(2)}\leqslant\cdots\leqslant r_{(T)}$，并令 q 为一个正整数，则 k 的两个估计定义为

275

$$k_p(q)=-\frac{1}{\ln(2)}\ln\left[\frac{-r_{(q)}+r_{(2q)}}{-r_{(2q)}+r_{(4q)}}\right], \tag{7.22}$$

$$k_h(q)=\frac{-1}{q}\sum_{i=1}^{q}\{\ln[-r_{(i)}]-\ln[-r_{(q+1)}]\}, \tag{7.23}$$

其中变量 (q) 用来强调这个估计依赖于 q．Hill 估计和 Pickands 估计之间的 q 的选择是不同的．

几个研究者已经研究过，但是还没有一般的对于可以利用的最好选择的结论. Dekkers 和 DeHaan(1989) 证明对样本容量 T，如果 q 在一个恰当的选择间隔上是增加的，则 $k_p(q)$是相合的. 另外，$\sqrt{q}[k_p(q)-k]$是渐近正态的，均值为 0，方差为 $k^2(2^{-2k+1}+1)/[2(2^{-k}-1)\ln(2)]^2$. Hill 估计仅仅对弗雷歇族分布是适用的，但是当它适用时，它比 Pickands 估计更有效. Goldie 和 Smith(1987) 证明$\sqrt{q}[k_h(q)-k]$是渐近正态的，均值为 0，方差为 k^2. 实际中，可以画出 Hill 估计 $k_h(q)$对 q 的图形，并且寻找一个使得估计是稳定的恰当的 q.

7.5.3 股票收益率的应用

我们对 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日的日对数收益率应用极值理论. 收益率是用百分比测量的，样本量为 9 190(即 T=9 190). 图 7-3 显示了当子区间的长度为 21 时，极端日对数收益率的时间图形，这近似对应于一个月. 从图中可以很清楚看到，1987 年 10 月的股市崩溃. 排除 1987 年的股市崩溃，极端日对数收益率的范围是 0.5%到 13%.

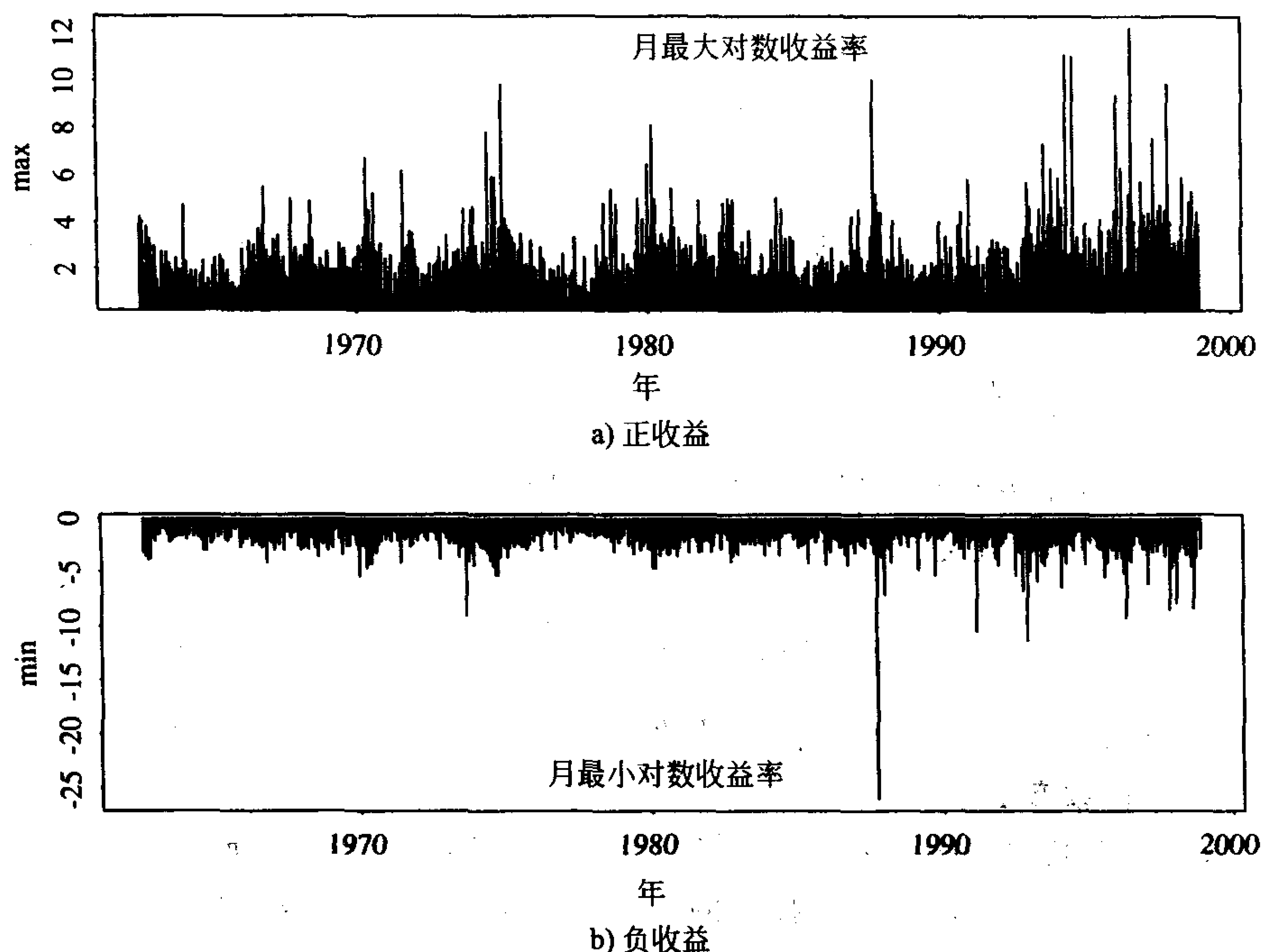


图 7-3 当子区间是 21 个交易日时，IBM 股票的最大与最小日对数收益率. 数据区间是从 1962 年 7 月 3 日至 1998 年 12 月 31 日

表 7-1 概括了由 Hill 估计得到的形状参数为 k 的一些估计结果. 此表显示对所给 q 的三种选择，结果都是稳定的. 为了对 Hill 估计的表现提供一个全面的概况，图 7-4 显示了 Hill 估计 $k_h(q)$对 q 的离散图. 对正的和负的日极值对数收益率，除了 q 很小的情形，估计都是稳定

的．估计的形状参数大约是−0.30，在渐近 5%水平下显著不同于 0．图形也说明了对于负收益极值，形状参数 k 看上去很小，隐含了日对数收益率可能有一个更厚的左尾．总之，结果说明 IBM 股票的日对数收益率的分布属于弗雷歇族．这样，该分析拒绝了实际中通常采用的正态性假定．这个结论与 Longin(1996)是一致的，他当时利用了美国股票市场指数序列．

表 7-1

q	190	200	210
最大值	−0.300(0.022)	−0.297(0.021)	−0.303(0.021)
最小值	−0.290(0.021)	−0.292(0.021)	−0.289(0.020)

注：IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日的日对数收益率 Hill 估计的结果，括号内的数为标准误差．

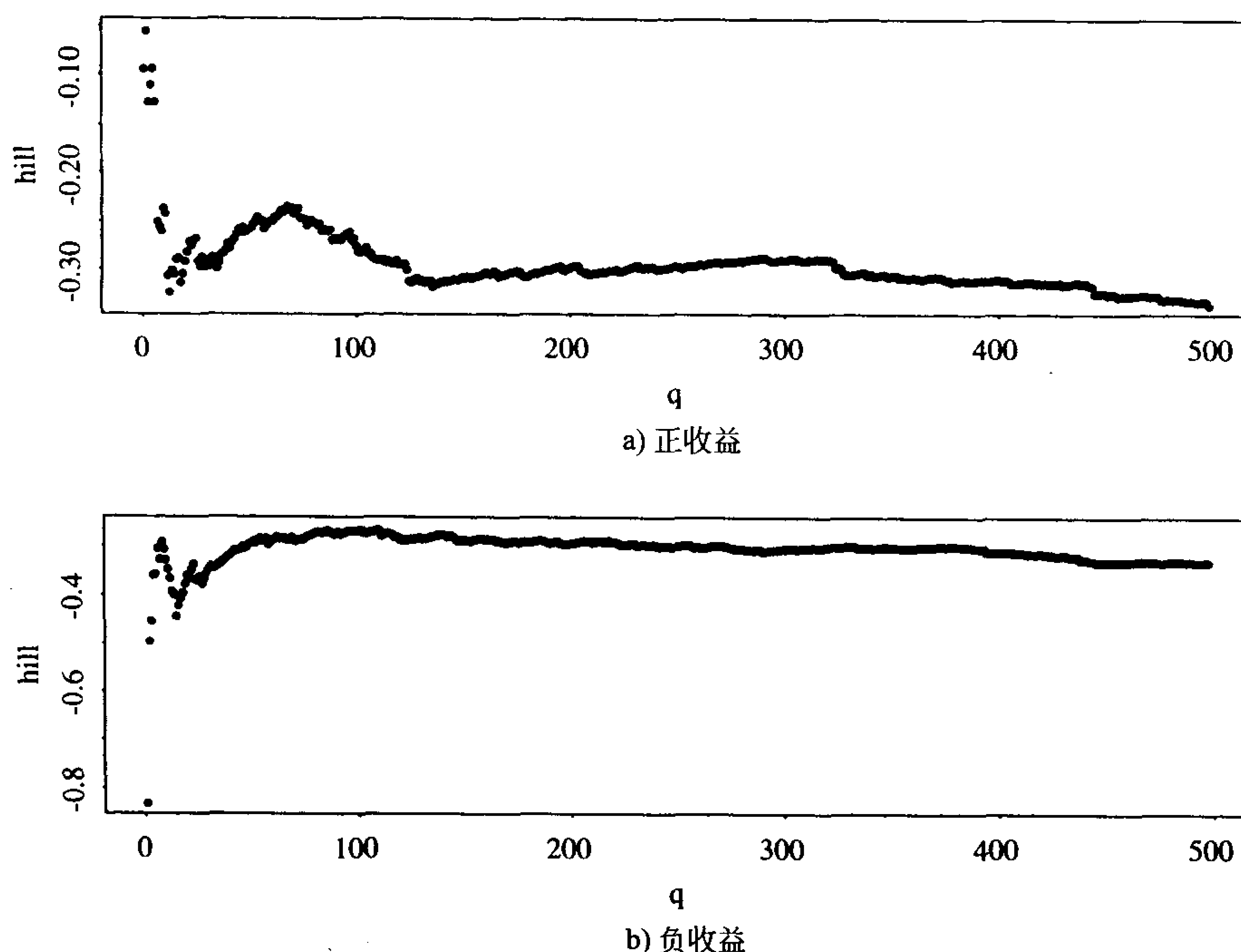


图 7-4 IBM 股票日对数收益率 Hill 估计的散点图．子样区间是从 1962 年 7 月 3 日至 1998 年 12 月 31 日

下面我们对 IBM 股票的日对数收益率应用最大似然法来估计一般极值分布的参数．表 7-2 对子区间长度从 1 个月(n=21)到 1 年(n=252)的变化范围的不同选择，概括了估计的结果．由此表，我们做出下面的观测结果：

- 当 n 增加时，位置参数 β_n 和尺度参数 α_n 的估计的绝对值是递增的．这是可以预料的，因为子区间最小值和最大值的绝对量值是 n 的非递减函数．
- 当 $n \geqslant 63$ 时，形状参数(或等价于尾指数)的估计对于负的极值是稳定的，近似为−0.33．
- 形状参数的估计对正极值不那么稳定．这个估计在数量上很小，但仍显著不同于 0．
- 因为 n=252 的子区间 g 的长度相当小，所以其结果具有高度可变性．

表 7-2

子区间长度	尺度 α_n	位置 β_n	形状参数 k_n
(a)最小收益			
1 个月($n=21$, $g=437$)	0.823(0.035)	−1.902(0.044)	−0.197(0.036)
1 季度($n=63$, $g=145$)	0.945(0.077)	−2.583(0.090)	−0.335(0.076)
6 个月($n=126$, $g=72$)	1.147(0.131)	−3.141(0.153)	−0.330(0.101)
1 年($n=252$, $g=36$)	1.542(0.242)	−3.761(0.285)	−0.322(0.127)
(b)最大收益			
1 个月($n=21$, $g=437$)	0.931(0.039)	2.184(0.050)	−0.168(0.036)
1 季度($n=63$, $g=145$)	1.157(0.087)	3.012(0.108)	−0.217(0.066)
6 个月($n=126$, $g=72$)	1.292(0.158)	3.471(0.181)	−0.349(0.130)
1 年($n=252$, $g=36$)	1.624(0.271)	4.475(0.325)	−0.264(0.186)

注：对 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日的日对数收益率极值分布的最大似然估计，括号内的数为标准误差.

276 ~ 278

得到的这个结果又类似于 Longin (1996)的结论，他对将极值理论应用到股票市场收益率的分析提供了一个很好的解释.

7.6 VaR 的极值方法

本节中，我们讨论利用极值理论的 VaR 计算方法. 这个方法类似于 Longin (1999a, 1999b)，他出于同样的目的提出了一个八步程序. 我们将讨论分为两个部分. 第一个部分关心的是利用前一小节中讨论的方法得到的参数估计；第二个部分通过将感兴趣的概率与不同的时间区间相联系，集中讨论 VaR 计算.

第一部分

假设样本区间上可以利用的资产收益率有 T 个观测值. 我们将样本区间分解为 g 个互不相交的长度为 n 的子区间，满足 $T=ng$. 如果 $T=ng+m$, $1\leqslant m<n$，那么我们将前 m 个观测值从样本中删除. 前一节中讨论的极值理论使我们能够得到对子区间的最小值 $\{r_{n,i}\}$ 的位置参数 β_n、尺度参数 α_n 以及形状参数 k_n 的估计. 将 MLE 估计代入方程(7.14)的 CDF, $x=(r-\beta_n)/\alpha_n$，我们可以得到一般极值分布在给定概率下的分位数. 因为我们集中于讨论持有多头金融头寸，所以感兴趣的是下尾概率(或左)分位数. 令 p^* 为一个小概率，它表示一个多头头寸的潜在损失超过一定限度的可能性，r_n^* 为子区间最小值在极限为一般极值分布条件下的 p^* 分位数，则我们有

$$p^*=\begin{cases}1-\exp\left[-\left(1+\dfrac{k_n(r_n^*-\beta_n)}{\alpha_n}\right)^{1/k_n}\right] & 若\ k_n\neq 0\\ 1-\exp\left[-\exp\left(\dfrac{r_n^*-\beta_n}{\alpha_n}\right)\right] & 若\ k_n=0\end{cases}$$

这里，对于 $k_n\neq 0$，有 $1+k_n(r_n^*-\beta_n)/\alpha_n>0$. 将这个方程改写为

$$\ln(1-p^*)=\begin{cases}-\left[1+\frac{k_n(r_n^*-\beta_n)}{\alpha_n}\right]^{1/k_n} & 若 k_n\neq 0\\ -\exp\left[\frac{r_n^*-\beta_n}{\alpha_n}\right] & 若 k_n=0\end{cases}$$

我们得到分位数为

$$r_n^*=\begin{cases}\beta_n-\frac{\alpha_n}{k_n}\{1-[-\ln(1-p^*)]^{k_n}\} & 若 k_n\neq 0\\ \beta_n+\alpha_n\ln[-\ln(1-p^*)] & 若 k_n=0\end{cases}\quad(7.24)$$

在金融应用中，主要感兴趣的就是 $k_n\neq 0$ 的情形.

279

第二部分

对于一个给定的下尾(或左尾)概率 p^*，方程(7.24)的分位数 r_n^* 就是对子区间最小值的极值理论基础上计算的 VaR. 下一步是制定子区间最小值与观测的收益率序列 r_t 之间的明确关系.

因为大部分资产收益率或者是序列无关的，或者有着很弱的序列相关性，我们可以利用方程(7.13)的关系，得到

$$p^*=P(r_{n,i}\leqslant r_n^*)=1-[1-P(r_t\leqslant r_n^*)]^n$$

或等价地

$$1-p^*=[1-P(r_t\leqslant r_n^*)]^n\quad(7.25)$$

概率间的这个关系允许我们得到原始的资产收益率序列的 VaR. 更精确地，对一个指定的很小的下尾概率 p，如果概率 p^* 是根据式(7.25)选择的，则 r_t 的 p 分位数是 r_n^*，这里 $p=P(r_t\leqslant r_n^*)$. 因此，对给定的小概率 p，持有一个对数收益率为 r_t 的标的资产的多头头寸的 VaR 为

$$\mathrm{VaR}=\begin{cases}\beta_n-\frac{\alpha_n}{k_n}\{1-[-n\ln(1-p)]^{k_n}\} & 若 k_n\neq 0\\ \beta_n+\alpha_n\ln[-n\ln(1-p)] & 若 k_n=0\end{cases}.\quad(7.26)$$

概括

我们将应用传统的极值理论来计算 VaR 的方法概括如下：

1)选择子区间的长度 n，并得到子区间的最小值 $\{r_{n,i}\}$，$i=1, \cdots, g$，这里 $g=T/n$.

2)得到 β_n、α_n 和 k_n 的最大似然估计.

3)检查拟合的极值模型的合适性，见下一节中模型检验的一些方法.

4)如果极值模型是充分的，则应用方程(7.26)计算 VaR.

注释： 因为我们集中讨论持有一个多头金融头寸，即讨论收益率分布的左尾分位数，因此，分位数是负的. 然而，实际中习惯于利用一个正数来计算 VaR，这样，在利用方程(7.26)时，应该意识到负号代表一个损失.

例 7.6　考虑 IBM 股票从 1962 年 7 月 7 日至 1998 年 12 月 31 日，以百分比表示的日对数收益. 由表 7-2，对 $n=63$，我们有 $\hat{\alpha}_n=0.945$，$\hat{\beta}_n=-2.583$，$\hat{k}_n=-0.335$. 因此，对左尾概

280

率 $p=0.01$，相应的 VaR 为

$$\text{VaR}=-2.583-\frac{0.945}{-0.335}\{1-[-63\ln(1-0.01)]^{-0.335}\}$$
$$=-3.04969$$

这样，对于股票的日收益率，1%分位数是−3.049 69. 如果某人持有这只股票价值1千万美元的一个多头头寸，则概率1%的 VaR 估计为 10 000 000×0.030 496 9=304 969 美元. 如果概率是0.05，则相应的 VaR 为 166 641 美元.

如果我们选择 $n=21$（即近似于1个月），那么$\hat{\alpha}_n=0.823$，$\hat{\beta}_n=-1.902$，$\hat{k}_n=-0.197$，极值分布的1%分位数为

$$\text{VaR}=-1.902-\frac{0.823}{-0.197}\{1-[-21\ln(1-0.01)]^{-0.197}\}=-3.40013$$

所以，对于一个1千万美元的多头头寸，在1%风险水平上对应的1天持有期的 VaR 为 340 013 美元. 如果概率是0.05，则对应的 VaR 为 184 127 美元. 在这个特殊的情形，$n=21$ 的选择给出了更高的 VaR 值.

例7.6中用极值理论得到的 VaR 值小于例7.3中用 GARCH(1，1)模型得到的 VaR 值，这一点有些奇怪. 事实上，例7.6中的 VaR 值甚至小于例7.5中根据经验分位数得到的 VaR 值. 出现这种情况的部分原因是由于概率为0.05的选择. 如果某人选取概率为 0.001=0.1%，并考虑同样的金融头寸，那么我们有 $n=21$ 时，对高斯 AR(2)-GARCH(1，1)模型，VaR=546 641 美元；对于极值理论有 VaR=666 590 美元. 而且，这里通过传统的极值理论得到的 VaR 可能并不充分，因为统计检验经常拒绝日对数收益率的独立性假定. 最后，子区间最小值的应用忽视了日对数收益率中的波动率聚类的事实. 下一节中要讨论的极值理论的新方法克服了这些缺点.

注释：例7.6中的结果显示，根据传统的极值理论计算的 VaR 依赖于 n 的选择，n 指子区间的长度. 因为极限的极值分布成立，所以选择一个大 n. 但是当样本量 T 固定时，一个大 n 意味着一个小 g，其中 g 为估计三个参数 β_n、α_n 和 k_n 时运用的有效样本量. 因此，需要在 n 与 g 的选择中作一些妥协. 应根据所研究资产的收益率给出一个恰当的选择. 我们建议在运用传统的极值理论时，应该检验结果中 VaR 的稳定性.

7.6.1 讨论

281

我们已经对一个1千万美元的多头头寸中 IBM 股票的日对数收益率的 VaR 计算运用了各种不同的方法. 考虑下一个交易日头寸的 VaR. 如果概率是5%，这意味着下一个交易日中损失将以概率0.95低于或等于 VaR，得到的结果为：

1)对于风险度量制为 302 500 美元；

2)对高斯 AR(2)-GARCH(1，1)模型为 287 200 美元；

3)对自由度为5的标准化学生-t 分布的 AR(2)-GARCH(1，1)模型为 283 520 美元；

4)运用经验分位数时，为 216 030 美元；

5)利用月最小值(即子区间长度 $n=21$)时，运用传统的极值理论方法，VaR 为 184 127 美元.

如果概率为 1%，则 VaR 为

1)对于风险度量制为 426 500 美元；

2)对高斯 AR(2)-GARCH(1，1)模型为 409 738 美元；

3)对自由度为 5 的标准化学生-t 分布的 AR(2)-GARCH(1，1)模型为 475 943 美元；

4)运用经验分位数时，为 365 709 美元；

5)利用月最小值(即子区间长度 $n=21$)时，运用传统的极值理论方法，VaR 为 340 013 美元.

如果概率为 0.1%，则 VaR 变为

1)对于风险度量制为 566 443 美元；

2)对高斯 AR(2)-GARCH(1，1)模型为 546 641 美元；

3)对自由度为 5 的标准化学生-t 分布的 AR(2)-GARCH(1，1)模型为 836 341 美元；

4)运用经验分位数时，为 780 712 美元；

5)利用月最小值(即子区间长度 $n=21$)时，运用传统的极值理论方法，VaR 为 666 590 美元.

不同方法间有着不小的区别. 这并不令人惊奇，因为在估计统计分布的尾行为时存在较大的不确定性. 由于没有真实的可以得到的 VaR 来比较不同方法的精度，我们建议运用几种方法来获得 VaR 的一个范围.

尾概率的选择在 VaR 计算中也起着重要的作用. 对于 IBM 股票的日收益率，因为样本量为 9 190，以至 5%和 1%的经验分位数都是收益率分布的分位数的适当估计. 在这种情形，我们可以将基于经验分位数的估计结果当作真实 VaR 的一个保守估计(即一个下界). 以这个观点，对 IBM 股票的日对数收益率而言，根据传统的极值理论方法看上去低估了 VaR. 下一节中讨论的条件极值理论的方法克服了这个缺点. [282]

当尾概率很小(如 0.1%)时，经验分位数是真实分位数的一个不太合理的估计. 基于经验分位数的 VaR 不能再充当真实 VaR 的一个下界. 最后，前面的结果很清楚地说明当尾概率很小时，在 VaR 计算中运用厚尾分布的效应. 在概率为 0.1%时，基于自由度为 5 的学生-t 分布的 VaR 或基于极值分布的 VaR 都高于基于正态假定下的 VaR.

7.6.2 多期 VaR

风险度量制方法的时间平方根法则变成了运用极值理论时的一个特殊情形. l 天持有期与 1 天持有期的关系为 $\text{VaR}(l)=l^{1/\alpha}\text{VaR}=l^{-k}\text{VaR}$，其中 α 是尾指数，k 是极值分布的形状参数；见 Danielsson 和 de Vries(1997a). 这个关系称为时间 α 根法则，这里 $\alpha=\frac{-1}{k}$，不是尺度参数 α_n.

为了解释，考虑例 7.6 中 IBM 股票的日对数收益率. 如果我们用 $p=0.05$ 以及 $n=21$ 的结果，则对 30 天持有期，我们有

$$\mathrm{VaR}(30) = (30)^{0.335}\mathrm{VaR} = 3.125 \times 184\ 127 = 575\ 397(\text{美元})$$

因为 $l^{0.335} < l^{0.5}$，则时间 α 根法则产生了比时间平方根法则更低的 l 天持有期的 VaR.

7.6.3 空头头寸的 VaR

在本小节中，我们给出持有一个空头头寸的 VaR 计算公式. 这里感兴趣的是子区间的最大值，并且极限的极值分布变为

$$F_*(r) = \begin{cases} \exp\left\{-\left[1 - \dfrac{k_n(r-\beta_n)}{\alpha_n}\right]^{1/k_n}\right\} & \text{若 } k_n \neq 0 \\ \exp\left[-\exp\left(\dfrac{r-\beta_n}{\alpha_n}\right)\right] & \text{若 } k_n = 0 \end{cases} \tag{7.27}$$

283 其中 r 表示子区间的最大值，这里对 $k_n \neq 0$，理解为有 $1 - k_n(r-\beta_n)/\alpha_n > 0$.

与多头头寸的程序类似，我们得到收益率 r_t 的 $(1-p)$ 分位数为

$$\mathrm{VaR} = \begin{cases} \beta_n + \dfrac{\alpha_n}{k_n}\{1 - [-n\ln(1-p)]^{k_n}\} & \text{若 } k_n \neq 0 \\ \beta_n + \alpha_n \ln[-n\ln(1-p)] & \text{若 } k_n = 0 \end{cases} \tag{7.28}$$

其中 p 是一个小概率，表示持有一个空头头寸时损失的机会.

7.7 基于极值理论的一个新方法

前面提到的利用极值理论计算 VaR 的方法遇到了一些困难. 首先，子区间长度 n 的选择并没有清楚地定义. 第二，该方法是无条件的，从而没有考虑其他解释变量的影响. 为了克服这些困难，统计文献中已经提出了极值理论的现代方法，见 Davison 和 Smith(1990)以及 Smith(1989). 新方法不是着重于讨论极值(最小值或最大值)，而是着重讨论股票的日对数收益率，对某个高门限的超出量和超出发生的时间. 例如，考虑 IBM 以及持有该股票的一个多头头寸. 令 η 表示一个指定的高门限，我们可以选择 $\eta = -2.5\%$. 假设第 i 次超越在第 t_i 天发生(即 $r_{t_i} \leqslant \eta$)，则新方法集中于讨论数据 $(t_i, r_{t_i} - \eta)$，$r_{t_i} - \eta$ 为超过门限 η 的超越量，t_i 表示第 i 次超越发生的时间. 类似地，对一个空头头寸，我们可以选择 $\eta = 2\%$，并讨论数据 $(t_i, r_{t_i} - \eta)$，这里 $r_{t_i} \geqslant \eta$.

实际中，发生时间 $\{t_i\}$ 对于重要的"稀少事件"(如对多头头寸而言低于门限 η)的出现强度提供了有用的信息. t_i 的一个聚类表示了一个大的市场低迷期. 超越量 $r_{t_i} - \eta$ 也是很重要的，因为它提供了我们所感兴趣的实际量.

根据前面的介绍，新方法并不要求子区间长度 n 的选择，但是它要求指定一个门限 η. 门限 η 的不同选择将导致形状参数 k(从而尾指数 $-1/k$)的不同估计. 在文献中，一些研究者相信 η 的选择既是一个统计问题，又是一个金融问题，它不能纯粹地根据统计理论来确定. 例如，不同的金融机构(或投资者)具有不同的风险容忍度. 这样，他们即使对于相同的金融头寸也可以选择不同的门限. 对于本章中考虑的 IBM 股票的日对数收益率，计算的 VaR 对 η 的选择是

不敏感的.

门限 η 的选择也依赖于观测到的对数收益率. 对一个稳定的收益率序列，选择 $\eta=-2.5\%$ 对多头头寸而言是相当好的. 对一个有较大波动的收益率序列(例如网络股的日收益率)，η 可能低至 -10%. 有限的经验显示，η 可以这样选择，使得超越的次数足够大(如大约为样本总数的 5%). 对于选择 η 的一个更加正式的研究，见 Danielsson 和 Vries (1997b). [284]

7.7.1 统计理论

再一次考虑资产的对数收益率 r_t. 假定第 i 个超越在 t_i 时刻发生. 集中于讨论超越量 $r_t-\eta$ 和超越时刻 t_i 导致统计想法的重要变化. 新方法不是使用边际分布(如最小值或最大值的极限分布)，而是采用一个条件分布来处理给定观测超过一个门限的条件下，超越量的大小. 超过门限的机会是由概率规律控制的. 换句话说，新方法考虑了多头头寸在给定 $r_t\leqslant\eta$，$x=r_t-\eta$ 的条件分布. 事件 $\{r_t\leqslant\eta\}$ 的出现服从一个点过程(例如一个泊松过程). 对泊松过程的定义见 6.9 节. 具体地，如果过程的强度参数 λ 是不随时间变的，则泊松过程是齐次的；如果 λ 是随时间变化的，则泊松过程是非齐次的. 泊松过程的概念可以一般化到多元情形.

为了表示方便，下面我们利用一个正门限以及收益率分布的右侧来讨论极值理论新方法背后的统计理论，这相当于持有一个金融空头寸. 然而，如果将此理论运用到 r_t^c 序列，$r_t^c=-r_t$，则该理论对于一个多头头寸适用得也同样好. 这一点很容易看出，因为对一个正门限 $r_t^c\geqslant\eta$，等价于 $r_t\leqslant-\eta$，这里 $-\eta$ 变为一个负门限.

新方法的基本理论是对于方程(7.27)中给出的最大值的极限分布考虑给定 $r>\eta$ 下 $r=x+\eta$ 的条件分布. 因为没有必要选择子区间的长度 n，所以我们不用它表示一个参数的下标. 给定 $r>\eta$ 条件下，$r\leqslant x+\eta$ 的条件分布为

$$\Pr(r\leqslant x+\eta\mid r>\eta)=\frac{\Pr(\eta\leqslant r\leqslant x+\eta)}{\Pr(r>\eta)}=\frac{\Pr(r\leqslant x+\eta)-\Pr(r\leqslant\eta)}{1-\Pr(r\leqslant\eta)}\tag{7.29}$$

利用方程(7.27)中的 CDF $F_*(.)$以及近似 $e^{-y}\approx1-y$，经过一些代数运算之后，我们得到

$$\begin{aligned}\Pr(r\leqslant x+\eta\mid r>\eta)&=\frac{F_*(x+\eta)-F_*(\eta)}{1-F_*(\eta)}\\&=\frac{\exp\left\{-\left[1-\dfrac{k(x+\eta-\beta)}{\alpha}\right]^{1/k}\right\}-\exp\left\{-\left[1-\dfrac{k(\eta-\beta)}{\alpha}\right]^{1/k}\right\}}{1-\exp\left\{-\left[1-\dfrac{k(\eta-\beta)}{\alpha}\right]^{1/k}\right\}}\\&\approx1-\left[1-\frac{kx}{\alpha-k(\eta-\beta)}\right]^{1/k}\end{aligned}\tag{7.30}$$

[285]

其中 $x>0$，且 $1-k(\eta-\beta)/\alpha>0$. 以后将会看出，这个逼近使得新方法与传统的极值理论之间的连接很明确了. $k=0$ 的情形看作是 $k\to0$ 的极限，从而

$$\Pr(r\leqslant x+\eta\mid r>\eta)\approx1-\exp(-x/\alpha).$$

7.7.2 一个新方法

利用方程(7.30)的统计结果，并且联合考虑超越量与超出时间，Smith (1989)提出了一个二维泊松过程来对(t_i, r_{t_i})建模. Tsay (1999)利用这个方法来研究风险管理中的VaR. 我们利用同样的方法.

假定基本的时间间隔为T，典型的取法是1年. 在美国，因为1年中一般有252个交易日，所以使用$T=252$. 令t表示数据点的时间间隔(例如日)，并将时间区间表示为$t=1, 2, \cdots, N$，这里N代表全部数据点的总个数. 对一个给定的门限η，超出门限的超出时间用$\{t_i, i=1, \cdots, N_\eta\}$来表示，在$t_i$点观测到的对数收益率为$r_{t_i}$. 因此，我们集中讨论对$\{(t_i, r_{t_i})\}$，$i=1, \cdots, N_\eta$建模，这里$N_\eta$依赖于门限$\eta$.

运用到极值理论的新方法猜想超越时间与相关收益率(即(t_i, r_{t_i}))联合形成一个二维的泊松过程，其强度测度由下式给出

$$\Lambda[(T_2, T_1)\times(r,\infty)] = \frac{T_2-T_1}{T}S(r;k,\alpha,\beta), \tag{7.31}$$

其中$S(r; k, \alpha, \beta)=\left[1-\frac{k(r-\beta)}{\alpha}\right]_+^{1/k}$，$0\leqslant T_1\leqslant T_2\leqslant N$，$r>\eta$，$\alpha>0$，$\beta$和$k$是参数，记号$[x]_+$定义为$[x]_+=\max(x, 0)$. 这个强度测度说明，超越门限的出现与时间段$[T_1, T_2]$的长度成正比，并且此概率可以由类似于方程(7.27)中CDF $F_*(r)$的指数的生存函数来控制. 随机变量X的生存函数定义为$S(x)=\Pr(X>x)=1-\Pr(X\leqslant x)=1-\text{CDF}(x)$. 当$k=0$时，强度测度可以看作$k\to 0$的极限，即

286

$$\Lambda[(T_2, T_1)\times(r,\infty)] = \frac{T_2-T_1}{T}\exp\left[\frac{-(r-\beta)}{\alpha}\right].$$

方程(7.31)中，时间间隔的长度是相对于基本区间长度T来度量的.

在考虑时间段$[0, T]$上，给定$r>\eta$条件下，$r=x+\eta(x>0)$隐含的条件概率时，利用方程(7.31)中的强度测度的思想变得非常清楚.

$$\frac{\Lambda[(0,T)\times(x+\eta,\infty)]}{\Lambda[(0,T)\times(\eta,\infty)]} = \left[\frac{1-k(x+\eta-\beta)/\alpha}{1-k(\eta-\beta)/\alpha}\right]^{1/k} = \left[1-\frac{kx}{\alpha-k(\eta-\beta)}\right]^{1/k}$$

恰是方程(7.30)中给出的条件分布的精确生存函数. 这个生存函数是由方程(7.27)给出的最大值的极限极值分布得到的. 这里我们利用生存函数，因为它表示超越发生的概率.

方程(7.27)中的极限极值分布与方程(7.31)中的强度测度的关系直接将极值理论的新方法与传统方法连接起来.

数学上，方程(7.31)的强度测度可以写为一个强度函数的积分

$$\Lambda[(T_2, T_1)\times(r,\infty)] = \int_{T_1}^{T_2}\int_r^{\infty}\lambda(t,z;k,\alpha,\beta)\,\mathrm{d}t\mathrm{d}z$$

其中强度函数$\lambda(t, z; k, \alpha, \beta)$定义为

$$\lambda(t,z;k,\alpha,\beta) = \frac{1}{T}g(z;k,\alpha,\beta), \tag{7.32}$$

这里

$$g(z;k,\alpha,\beta)=\begin{cases}\dfrac{1}{\alpha}\left[1-\dfrac{k(z-\beta)}{\alpha}\right]^{1/k-1} & 若\ k\neq 0\\ \dfrac{1}{\alpha}\exp\left[\dfrac{-(z-\beta)}{\alpha}\right] & 若\ k=0\end{cases}$$

利用泊松过程的结果，我们可以写下观测到的超越时间与相应收益率$\{(t_i, r_{t_i})\}$在二维空间$[0, N]\times(\eta, \infty)$上的似然函数为

$$L(k,\alpha,\beta)=\left(\prod_{i=1}^{N_\eta}\frac{1}{T}g(r_{t_i};k,\alpha,\beta)\right)\times\exp\left[-\frac{N}{T}S(\eta;k,\alpha,\beta)\right], \tag{7.33}$$

参数 k，α，β 可以通过最大化这个似然函数的对数来估计. 因为尺度参数 α 非负，所以在估计中我们利用 $\ln(\alpha)$.

287

例 7.7 再次考虑 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日的日对数收益率，共有 9 190 个日收益率的值. 表 7-3 给出了负序列$\{-r_t\}$在门限的三种选择下，参数 k，α，β 的一些估计结果. 我们利用负序列$\{-r_t\}$，而不是$\{r_t\}$，因为集中于讨论持有一个多头的金融头寸. 此表也给出了给定门限下超越次数. 可以看出 IBM 股票在 1 天中降低 2.5%或者更多的机会出现的概率为 310/9 190≈3.4%. 因为 IBM 股票收益率的样本均值不是 0，所以我们也考虑了当样本均值从原始的日对数收益率中去除以后的情况. 由表中看出，去除样本均值后对参数估计只有很小的影响. 下面利用这些参数估计来计算 VaR，记住在实际应用中必须要仔细检查拟合一个泊松模型的合适性. 下一小节中我们将讨论模型检验的方法.

表 7-3

Thr.	Exc.	形状参数 k.	对数(尺度)$\ln(\alpha)$	位置 β
		(a)原始的对数收益		
3.0%	175	−0.306 97(0.090 15)	0.306 99(0.123 80)	4.692 04(0.190 58)
2.5%	310	−0.264 18(0.065 01)	0.315 29(0.112 77)	4.740 62(0.180 41)
2.0%	554	−0.187 51(0.043 94)	0.276 55(0.098 67)	4.810 03(0.172 09)
		(b)移除样本均值		
3.0%	184	−0.305 16(0.088 24)	0.308 07(0.123 95)	4.738 04(0.191 51)
2.5%	334	−0.281 79(0.067 37)	0.319 68(0.120 65)	4.768 08(0.185 33)
2.0%	590	−0.192 60(0.043 57)	0.279 17(0.099 13)	4.848 59(0.172 55)

注：对 IBM 股票从 1962 年 7 月 3 日到 1998 年 12 月 31 日的负日对数收益率的二维齐次泊松模型的估计结果. 基本的时间间隔为 252 天(即一年). 括号内的数为标准误差，其中“Thr.”和“Exc”表示门限和超越次数.

7.7.3 基于新方法的 VaR 计算

正如方程(7.30)显示的，运用的二维泊松过程模型(它采用了方程(7.31)中的强度测度)与方程(7.27)中的极值分布具有同样的参数值. 因此，可以利用与方程(7.28)同样的公式计算新方法下的 VaR. 更具体地，对一个给定的上尾概率 p，对数收益率 r_t 的$(1-p)$分位数为

$$\text{VaR}=\begin{cases}\beta+\dfrac{\alpha}{k}\{1-[-T\ln(1-p)]^{k}\} & \text{若 } k\neq 0\\ \beta+\alpha\ln[-T\ln(1-p)] & \text{若 } k=0\end{cases}\tag{7.34}$$

其中 T 为估计中使用的基本时间间隔．典型地，美国利用 $T=252$ 表示 1 年中交易日的近似

288

数量．

例 7.8　再次考虑持有一个价值 1 千万美元的 IBM 股票的多头头寸的情形．我们利用表 7-3 中的估计结果对尾概率为 0.05 和 0.01 计算持有期为 1 天的 VaR.

- 情形 1：利用原始的日对数收益率．门限 η 的三种选择导致了下面的 VaR 值：

1) $\eta=3.0\%$：VaR(5%)=228 239(美元)，　VaR(1%)=359 303(美元)；

2) $\eta=2.5\%$：VaR(5%)=219 106(美元)，　VaR(1%)=361 119(美元)；

3) $\eta=2.0\%$：VaR(5%)=212 981(美元)，　VaR(1%)=368 552(美元)；

- 情形 2：消除日对数收益率的样本均值．门限 η 的三种选择导致了下面的 VaR 值：

1) $\eta=3.0\%$：VaR(5%)=232 094(美元)，　VaR(1%)=363 697(美元)；

2) $\eta=2.5\%$：VaR(5%)=225 782(美元)，　VaR(1%)=364 254(美元)；

3) $\eta=2.0\%$：VaR(5%)=217 740(美元)，　VaR(1%)=372 372(美元)．

正如料想的，消除样本均值(正值)以后，稍微地增加了 VaR．然而，VaR 在门限的三种选择下是相当稳定的．实际中，我们建议将新方法运用到 VaR 计算中之前首先消除样本均值．

讨论：与例 7.6 中的 VaR 比较(使用了传统的极值理论)，新方法提供了一个更加稳定的 VaR 计算．传统方法对于子区间长度 n 的选择是相当敏感的．

7.7.4　解释变量的使用

前面讨论的二维泊松过程模型是齐次的，因为三个参数 k，α 和 β 是随时间不变的常数．实际中，这种模型可能并不合适．而且，通常可以利用能够影响对数收益率 r_t 的一些解释变量．新的极值理论方法对 VaR 计算的一个良好性质是能很容易地考虑解释变量．在本小节中，我们将讨论这样一个框架．另外，我们也讨论可以用来检验一个拟合的二维泊松过程模型合适性的方法．

假定 $\boldsymbol{x}_t=(x_{1t},\cdots,x_{vt})'$ 是 v 个解释变量的向量，在时刻 t 以前就可以得到．对资产收益率，第 3 章讨论的 r_t 的波动率 σ_t^2 是一个解释变量的例子．美国证券市场上解释变量的另一个

289

例子是表示联邦公开市场委员会(Federal Open Market Committee)会议的示性变量．利用解释变量的一个简单方法是猜想三个参数 k，α 和 β 是时变的，而且是解释变量的线性函数．具体地，当解释变量 $\boldsymbol{x}_t$ 可以得到时，我们假定

$$\begin{aligned}k_t&=\gamma_0+\gamma_1x_{1t}+\cdots+\gamma_vx_{vt}\equiv\gamma_0+\boldsymbol{\gamma}'\boldsymbol{x}_t\\ \ln(\alpha_t)&=\delta_0+\delta_1x_{1t}+\cdots+\delta_vx_{vt}\equiv\delta_0+\boldsymbol{\delta}'\boldsymbol{x}_t\\ \beta_t&=\theta_0+\theta_1x_{1t}+\cdots+\theta_vx_{vt}\equiv\theta_0+\boldsymbol{\theta}'\boldsymbol{x}_t\end{aligned}\tag{7.35}$$

如果 $\boldsymbol{\gamma}=\boldsymbol{0}$，则形状参数 $k_t=\gamma_0$，它是不随时间变化的．这样，检验 γ 的显著性可以提供关于解释变量对形状参数的贡献的信息．类似的方法可以运用到尺度参数和位置参数．在方程(7.35)

中，我们对所有的三个参数 k_t，$\ln(\alpha_t)$和 β_t 都使用同样的解释变量．在应用时，不同的参数可以使用不同的解释变量．

当极值分布的三个参数随时间变化的时候，我们得到一个非齐次的泊松过程．强度测度变为

$$\Lambda[(T_1,T_2)\times(r,\infty)]=\frac{T_2-T_1}{T}\left(1-\frac{k_t(r-\beta_t)}{\alpha_t}\right)_+^{1/k_t},\qquad r>\eta. \tag{7.36}$$

超出时间点和收益率$\{(t_i,\ r_{t_i})\}$的似然函数变为

$$L=\left(\prod_{i=1}^{N_\eta}\frac{1}{T}g(r_{t_i};k_{t_i},\alpha_{t_i},\beta_{t_i})\right)\times\exp\left[-\frac{1}{T}\int_0^N S(\eta;k_t,\alpha_t,\beta_t)dt\right],$$

如果假定参数 k_t，α_t 和 β_t 在每个交易日内是固定的，其中 $g(z;\ k_t,\ \alpha_t,\ \beta_t)$和 $S(\eta;\ k_t,\ \alpha_t,\ \beta_t)$分别由方程(7.32)和方程(7.31)给出，则似然函数简化为

$$L=\left(\prod_{i=1}^{N_\eta}\frac{1}{T}g(r_{t_i};k_{t_i},\alpha_{t_i},\beta_{t_i})\right)\times\exp\left[-\frac{1}{T}\sum_{t=1}^{N}S(\eta;k_t,\alpha_t,\beta_t)\right]. \tag{7.37}$$

对给定的观测值$\{r_t,\ \boldsymbol{x}_t\mid t=1,\ \cdots,\ N\}$，基本的时间间隔为 T，门限为 η，方程(7.35)的参数可以通过最大化方程(7.37)似然函数的对数来估计．这里又一次使用 $\ln(\alpha_t)$来满足 α_t 为正的限制．

注释： 从三个参数都是时刻 t 时可以得到信息的精确函数这个意义上讲，方程(7.35)的参数化类似于第 3 章中波动率模型的参数化．如果有必要也可以使用其他函数．

7.7.5 模型检验

对超越次数和超出量(excesses)检验的二维泊松过程模型进行检验涉及到检验模型的三个重要性质．第一个性质是证明超越率的合适性，第二个性质是检验超越的分布，最后一个性质是检验模型的独立性假设．我们主要讨论在检验这三个性质时有用的一些统计量．这些统计量基于一些基本的、与分布和随机过程相关的基本统计理论．　290

超越率

一元泊松过程的一个基本性质是接连两个事件之间的时间持续期是独立的，且都服从指数分布．为了采用对检验一个二维过程模型的类似的性质，Smith 和 Shively (1995)提出研究接连两个超越间的时间持续期．如果这个二维的泊松过程模型对超越次数和超越量是恰当的，则第 i 个超越与第 $i-1$ 个超越之间的时间持续期应该服从一个指数分布．更具体地，令 $t_0=0$，我们期望

$$z_{t_i}=\int_{t_{i-1}}^{t_i}\frac{1}{T}g(\eta;k_s,\alpha_s,\beta_s)\mathrm{d}s,\qquad i=1,2,\cdots,$$

是独立同分布(iid)的，服从一个标准指数分布．因为日收益率是离散时间的观测值，我们采用的时间持续期为

$$z_{t_i}=\frac{1}{T}\sum_{t=t_{i-1}+1}^{t_i}S(\eta;k_t,\alpha_t,\beta_t), \tag{7.38}$$

并利用分位数-分位数(QQ)图来检验 iid 标准指数分布的有效性．如果模型是正确的，则 QQ

图应显示一条通过原点且斜率为 1 的直线.

超越量的分布

在所考虑的二维泊松过程模型下，超过门限 η 的超出量 $x_t = r_t - \eta$ 的条件分布是一个形状参数为 k_t，尺度参数为 $\psi_t = \alpha_t - k_t(\eta - \beta_t)$ 的广义帕雷托分布(GPD). 因此，我们可以利用标准指数分布与 GPD 之间的关系，定义

$$w_{t_i} = \begin{cases} \dfrac{-1}{k_{t_i}} \ln\left(1 - k_{t_i} \dfrac{r_{t_i} - \eta}{\psi_{t_i}}\right)_+ & 若 k_{t_i} \neq 0 \\ \dfrac{r_{t_i} - \eta}{\psi_{t_i}} & 若 k_{t_i} = 0 \end{cases}, \tag{7.39}$$

如果模型是充分的，则 $\{w_{t_i}\}$ 是独立的均值为 1 的指数分布，见 Smith(1999). 我们可以利用 QQ 图对检验超出量的 GPD 假定的有效性.

291

独立性

在调整解释变量的影响以后，检验独立性假定的一个方法是研究 z_{t_i} 与 w_{t_i} 的样本自相关函数. 在独立性假定下，我们期望 z_{t_i} 和 w_{t_i} 的序列相关系数都是零.

7.7.6 解释

在本小节中，我们对 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日以百分比表示的日对数收益率采用一个二维非齐次的泊松过程模型. 我们集中于讨论持有 1 千万美元的多头头寸. 该分析使我们能够与前面利用其他方法计算的 VaR 得到的结果相比较.

首先我们指出例 7.7 中的二维齐次模型需要进一步提炼，因为拟合的模型没能通过前一小节中的模型统计量的检验. 图 7-5a、b 显示了齐次模型当门限为 $\eta = 2.5\%$ 时，z_{t_i} 和 w_{t_i} 的自相关函数，其定义分别见方程(7.38)与方程(7.39). 图形中的两条水平线表示两个标准误差的渐近上下限. 可以看出 z_{t_i} 和 w_{t_i} 序列都有某些显著的序列相关. 图 7-6a、b 显示了 z_{t_i} 和 w_{t_i} 序列的 QQ 图，每个图形中的直线都为理论线，它过原点并且在标准指数分布假定下斜率为 1. z_{t_i} 的 QQ 图显示了有些不相符.

为了提炼模型，我们使用均值修正的对数收益率序列

$$r_t^o = r_t - \bar{r}, \quad \bar{r} = \frac{1}{9\,190} \sum_{t=1}^{9\,190} r_t,$$

其中 r_t 为百分比表示的日对数收益率，并且采用下面的解释变量:

1) x_{1t}: 对 10 月、11 月和 12 月的示性变量. 即，如果 t 在 10 月、11 月或 12 月，则 $x_{1t} = 1$. 选择这个变量是考虑到第四季度对 IBM 日收益率的影响(或称年末效应)，如果这种效应存在的话.

2) x_{2t}: 对前一个交易日行为的示性变量. 具体地，$x_{2t} = 1$ 当且仅当对数收益率 $r_{t-1}^o \leqslant -2.5\%$. 因为我们集中于讨论门限为 2.5% 的一个多头头寸，当价格降低超过 2.5% 时超越发生. 因此，当 IBM 股票的价格相对于前一天交易日的价格降低 2.5% 或更多时，使用 x_{2t} 来捕获恐慌抛售的可能性.

3) x_{3t}：波动率的一个定性测度，它是 $t-1$ 与 $t-5$（包含 $t-1$ 和 $t-5$）之间的具有超过门限的对数收益率的天数. 在所考虑的情形，x_{3t}是满足 $|r^o_{t-i}|\geqslant 2.5\%$，$i=1$，…，5 的 r^o_{t-i}的个数.

4) x_{4t}：年趋势，定义为 $x_{4t}=$（时刻 t 所在的年份 $-1961)/38$. 用此变量来识别 IBM 股票极值收益率的行为中的任何趋势.

5) x_{5t}：基于对均值修正序列 r^o_t 的高斯 GARCH(1, 1)模型的波动率序列. 具体地，$x_{5t}=\sigma_t$，这里 σ^2_t为 GARCH(1, 1)模型的条件方差.

$$r^o_t = a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1)$$
$$\sigma^2_t = 0.045\ 65 + 0.080\ 7a^2_{t-1} + 0.903\ 1\sigma^2_{t-1},$$

这 5 个解释变量在时刻 $t-1$ 都可以得到. 我们利用两个波动率测度（x_{3t}与 x_{5t}）来研究市场波动率对 VaR 的影响. 正如例 7.3 中拟合的 AR(2)-GARCH(1, 1)模型显示的，r_t 中的序列相关性是很弱的，以至于我们不能对均值方程接受任何 ARMA 模型.

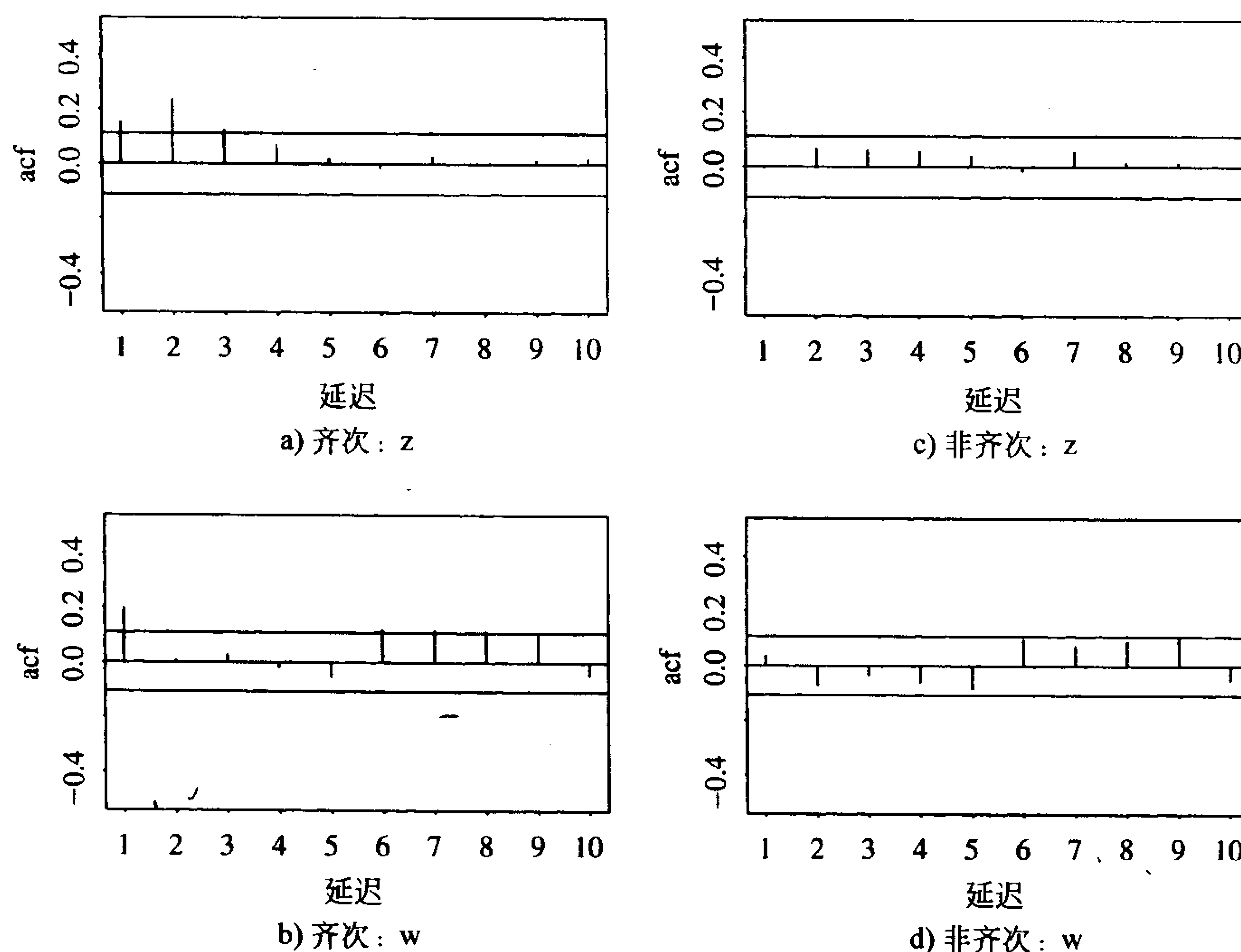


图 7-5 二维泊松过程模型中的 z 和 w 序列的样本自相关函数. a)与 b)显示对齐次模型的结果；c)与 d)显示对非齐次模型的结果. 所用数据为 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日以百分比表示的、经均值修正后的日对数收益率，门限为 $\eta=2.5\%$. 所使用的是多头金融头寸

利用前面的 5 个解释变量，消除不显著的参数，我们得到表 7-4 显示的估计结果. 图 7-5c、d 以及图 7-6c、d 显示了当门限是 $\eta=2.5\%$时，拟合的二维非齐次泊松过程模型的模型检验统计量. z_{t_i}与 w_{t_i} 的所有自相关函数都在两个标准误差的渐近上下限之内. QQ 图也说明了显著的改善，因为它们表明没有模型的合适性. 根据检验结果，非齐次模型看上去是合适的.

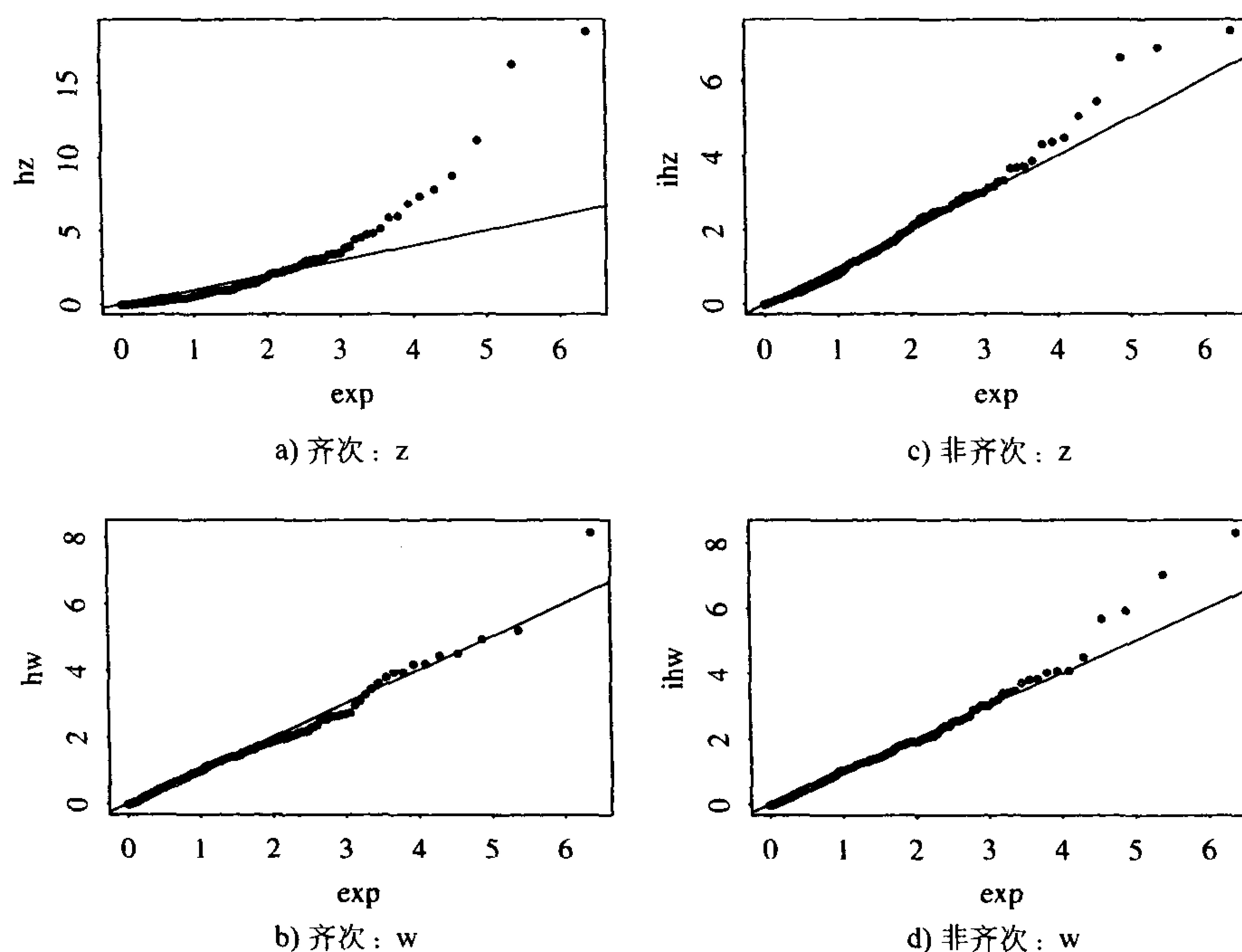


图 7-6 二维泊松过程模型中的 z 和 w 度量的 QQ 图. a)、b)显示对齐次模型的结果；c)与 d)显示对非齐次模型的结果. 所用数据为 IBM 股票从 1962 年 7 月 3 日至 1998 年 12 月 31 日以百分比表示的、经均值修正后的日对数收益率，门限为 η=2.5%. 所使用的是多头金融头寸

表 7-4

参数	常数	x_{3t}的系数	x_{4t}的系数	x_{5t}的系数
		(a)门限 2.5%，带 334 个超越		
β_t	0.320 2		1.477 2	2.199 1
(标准误差)	(0.338 7)		(0.322 2)	(0.245 0)
$\ln(\alpha_t)$	−0.811 9	0.330 5	1.032 4	
(标准误差)	(0.179 8)	(0.082 6)	(0.261 9)	
k_t	−0.180 5	−0.211 8	−0.355 1	0.260 2
(标准误差)	(0.129 0)	(0.058 0)	(0.150 3)	(0.046 1)
		(b)门限 3.0%，带 184 个超越		
β_t	1.156 9			2.191 8
(标准误差)	(0.408 2)			(0.290 9)
$\ln(\alpha_t)$	−0.031 6	0.333 6		
(标准误差)	(0.120 1)	(0.086 1)		
k_t	−0.600 8	−0.248 0		0.317 5
(标准误差)	(0.145 4)	(0.073 1)		(0.068 5)

注：对 IBM 股票从 1962 年 7 月 3 日到 1998 年 12 月 31 日以百分比表示的日对数收益的二维非齐次泊松过程模型的估计结果. 利用文中定义的四个解释变量. 此模型是为持有 IBM 股票的多头头寸设定的，消除了对数收益率数据的样本均值.

考虑门限为 2.5%的情形. 估计结果说明:

1)强度函数的所有三个参数都显著依赖于年时间趋势. 特别地，形状参数有一个负的年趋势，表明随时间的变化，IBM 股票的对数收益率偏离正态性越来越远. 位置参数和尺度参数都随时间增大.

2)第 4 季度的示性变量 x_{1t} 与恐慌抛售的示性变量 x_{2t} 对所有的三个参数都是不显著的.

3)由 GARCH(1，1)模型的波动率对位置和形状参数的影响是正的；见 x_{5t} 的系数. 这是可以理解的，因为当波动率高时，对数收益率的可变性增加了，从而降低了对数收益率对尾指数的依赖性.

4)尺度和形状参数显著依赖于波动率的定性度量. 估计的符号也是合理的.

1998 年 12 月 31 日的解释变量假定为 $x_{39\,190}=0$，$x_{49\,190}=0.973\,7$，$x_{59\,190}=1.976\,6$. 利用这些值以及表 7-4 中拟合的模型，假定尾概率为 0.05，我们得到

$$k_{9\,190}=-0.011\,95,\quad \ln(\alpha_{9\,190})=0.193\,31,\quad \beta_{9\,190}=6.105.$$

方程(7.34)显示的 VaR 分位数给出 VaR=3.037 56%. 因此，对一个 1 千万美元的多头头寸，我们有

$$\text{VaR}=10\,000\,000\times 0.030\,375\,6=303\,756(\text{美元})$$

如果尾概率是 0.01，则 VaR 为 497 425 美元. 5% 的 VaR 稍高于例 7.3 中利用高斯 AR(2)-GARCH(1，1)模型给出的结果. 1%的 VaR 高于例 7.3 情形 1 的结果. 如我们所料，极值(即厚尾)对 VaR 的影响当使用的尾概率较小时更加重要.

292
~
295

采用解释变量的一个优势是参数适应市场条件的变化. 例如，对 1998 年 12 月 30 日的解释变量假定的值为 $x_{39\,189}=1$，$x_{49\,189}=0.973\,7$，$x_{59\,189}=1.875\,7$. 在这种情形，我们有

$$k_{9\,189}=-0.250\,0,\quad \ln(\alpha_{9\,189})=0.523\,85,\quad \beta_{9\,189}=5.883\,4.$$

则 95% 分位数(即尾概率为 5%)变为 2.691 39%. 因此，VaR 变为

$$\text{VaR}=10\,000\,000\times 0.026\,913\,9=269\,139(\text{美元})$$

如果尾概率为 0.01，则 VaR 变为 448 323. 根据这个例子，例 7.8 中显示的齐次泊松模型似乎低估了 VaR.

练习题

1. 考虑 GE 股票从 1962 年 7 月 3 日至 1999 年 12 月 31 日的日对数收益率，此数据可以从 CRSP 或者文件“d-geln. dat”中得到. 假定你持有这只股票价值 100 万美元的多头头寸. 采用尾概率 0.05，并利用下面的方法计算 1 天持有期和 15 天持有期的风险值(VaR).

   (a)风险度量制方法.

   (b)高斯 ARMA-GARCH 模型.

   (c)带学生-t 分布的 ARMA-GARCH 模型，这时你也需要估计自由度.

   (d)子区间长度为 $n=21$ 的传统极值理论.

2. 文件“d. csco9199. dat”含有 Cisco Systems 股票从 1991 年到 1999 年的日对数收益率，共

2 275 个观测值. 假定你持有 Cisco Systems 股票的价值 100 万美元的多头头寸. 利用概率 $p=0.01$,计算你的头寸在下一个交易日的风险值.

(a)利用风险度量制方法.

(b)利用带条件高斯分布的 GARCH 模型.

(c)利用带学生-t 分布的 GARCH 模型，你也可能要估计自由度.

(d)利用无条件样本分位数.

(e)利用一个门限为 2%的二维齐次泊松过程. 也就是说，讨论股票日价格降低 2%或更多的超越次数和超越量. 检验拟合的模型.

(f)利用门限为 2%的二维非齐次泊松过程. 解释变量为：(1)年时间趋势，(2)对 10 月、11 月和 12 月的哑变量，(3)基于一个高斯 GARCH(1，1)模型拟合的波动率. 对拟合的模型实施诊断检验.

296

(g)利用门限 2.5%或 3%重复前面的二维非齐次泊松过程. 评论门限的选择.

3. 对 Cisco Systems 的日股票收益率利用 Hill 估计以及数据"d-csco9199.dat"估计尾指数.

4. 文件"d-hwp3dx8099.dat"含有惠普、CRSP 价值加权指数、等权重指数和标准普尔 500 指数从 1980 年到 1999 年的日对数收益率. 所有的收益率都以百分比表示，并且包含红利分布. 假定所感兴趣的尾概率为 0.01. 对下面的金融头寸计算 2000 年第一个交易日的风险值.

(a)持有 100 万美元的惠普股票和 100 万美元标准普尔 500 指数的多头，利用风险度量制方法. 对每个序列都需要估计 IGARCH(1，1)模型的 α 系数.

(b)与(a)部分同样的头寸，但是对每个收益率序列利用一个一元 ARMA-GARCH 模型.

(c)对持有 100 万美元的惠普股票的多头，利用一个二维非齐次泊松模型，解释变量如下：(1)年时间趋势；(2)对惠普股票基于高斯 GARCH 模型拟合的波动率，(3)对标准普尔 500 指数收益率基于高斯 GARCH 模型拟合的波动率；(4)对价值加权指数收益基于高斯 GARCH 模型拟合的波动率. 对拟合的模型实施诊断检验. 由标准普尔 500 指数与价值加权指数收益率测量的市场波动率在决定惠普的股票收益率的尾行为上有意义吗？你可以选择几种不同的门限.

参考文献

Berman, S. M. (1964), "Limiting theorems for the maximum term in stationary sequences," *Annals of Mathematical Statistics*, 35, 502–516.

Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

Danielsson, J., and De Vries, C. G. (1997a), "Value at risk and extreme returns," working paper, London School of Economics, London, U.K.

Danielsson, J., and De Vries, C. G. (1997b), "Tail index and quantile estimation with very high frequency data," *Journal of Empirical Finance*, 4, 241–257.

Davison, A. C., and Smith, R. L. (1990), "Models for exceedances over high thresholds," (with discussion), *Journal of the Royal Statistical Society, Series B*, 52, 393–442.

De Haan L., Resnick, I. S., Rootzén, and De Vries, C. G. (1989), "Extremal behavior of solutions to a stochastic difference equation with applications to ARCH process," *Stochastic Processes and Their Applications*, 32, 213–224.

Dekkers, A. L. M., and De Haan, L. (1989), "On the estimation of extreme value index and large quantile estimation," *Annals of Statistics*, 17, 1795–1832. [297]

Duffie, D., and Pan, J. (1997), "An overview of value at risk," *Journal of Derivatives*, Spring, 7–48.

Embrechts, P., Kuppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events*, Berlin: Springer Verlag.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, New York: Wiley.

Goldie, C. M., and Smith, R. L. (1987), "Slow variation with remainder: Theory and applications," *Quarterly Journal of Mathematics*, Oxford 2nd series, 38, 45–71.

Gnedenko, B. V. (1943), "Sur la distribution limite du terme maximum of d'une série Aléatorie," *Annals of Mathematics*, **44**, 423–453.

Gumbel, E. J. (1958), *Statistics of Extremes*, New York: Columbia University Press.

Hill, B. M. (1975), "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, 3, 1163–1173.

Jenkinson, A. F. (1955), "The frequency distribution of the annual maximum (or minimum) of meteorological elements," *Quarterly Journal of the Royal Meteorological Society*, **81**, 158–171.

Jorion, P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*. The McGraw-Hill Company: Chicago.

Koenker, R. W., and Bassett, G. W. (1978), "Regression quantiles," *Econometrica*, 46, 33–50.

Koenker, R. W., and D'Orey, V. (1987), "Computing regression quantiles," *Applied Statistics*, 36, 383–393.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer Verlag.

Longerstaey, J., and More, L. (1995), "Introduction to RiskMetrics^{TM}," 4th edition, Morgan Guaranty Trust Company: New York.

Longin, F. M. (1996), "The asymptotic distribution of extreme stock market returns," *Journal of Business*, 69, 383–408.

Longin, F. M. (1999a), "Optimal margin level in futures markets: Extreme price movements," *The Journal of Futures Markets*, 19, 127–152.

Longin, F. M. (1999b), "From value at risk to stress testing: The extreme value approach," working paper, Centre for Economic Policy Research, London, UK.

Pickands, J. (1975), "Statistical inference using extreme order statistics," *Annals of Statistics*, 3, 119–131.

Smith, R. L. (1989), "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone" (with discussion), *Statistical Science*, 4, 367–393.

Smith, R. L. (1999), "Measuring risk with extreme value theory," working paper, Department of Statistics, University of North Carolina at Chapel Hill.

Smith, R. L., and Shively, T. S. (1995), "A point process approach to modeling trends in troposheric ozone," *Atmospheric Environment*, **29**, 3489–3499.

Tsay, R. S. (1999), "Extreme value analysis of financial data," working paper, Graduate School of Business, University of Chicago. [298]

第8章 多元时间序列分析及其应用

近年来，经济全球化与互联网交流已经加速了世界金融市场的一体化. 单个市场的价格运动能够很容易且迅速地扩散到另一个市场，由于这个原因，金融市场比以前更加彼此依赖. 为了更好地理解全球金融的动态结构，必须将它们联合考虑. 在某些情况下，一个市场可能引导另一个市场，然而，这个关系在其他情形下可能颠倒过来. 因此，了解市场是如何相互关联的在金融中相当重要. 类似地，对一个持有多个资产的投资者或一个金融机构，资产收益率间的动态关系在决策制定中起着很重要的作用. 在本章和下一章中，我们引进有用的金融计量经济学模型和方法来联合研究多期收益序列. 在统计文献中，这些模型和方法属于向量或多元时间序列分析.

多元时间序列包含多个一元时间序列作为其分量. 因为这样，向量与矩阵的概念在多元时间序列分析中是非常重要的. 我们用黑体记号来表示向量和矩阵. 如果有必要，读者可以查阅本章中的附录 A 了解向量和矩阵的一些基本运算和性质. 附录 B 给出了多元正态分布的一些结果，这些结果在多元统计分析中广泛使用(例 Johnson 和 Wichern 1998).

令 $\boldsymbol{r}_t=(r_{1t}, r_{2t}, \cdots, r_{kt})'$表示$k$ 个资产在时刻t 的对数收益率，这里$\boldsymbol{a}'$表示$\boldsymbol{a}$ 的转置. 例如，一个持有 IBM、Microsoft、Exxon Mobil、General Motors 和 Wal-Mart Stores 股票的投资者可能考虑这些公司日对数收益率构成的 5 维向量. 这里 r_{1t}表示 IBM 股票的日对数收益，r_{2t}表示 Microsoft 的，等等. 作为第二个例子，考虑一个对全球投资感兴趣的投资者，他可能考虑美国标准普尔 500 指数、英国 FTSE100 指数以及日本的 Nikkei225 指数的收益率序列. 这时序列是 3 维的，r_{1t}表示标准普尔 500 指数的收益率，r_{2t}表示 FTSE100 指数的收益率，r_{3t}表示 Nikkei225 指数的收益率. 本章的目的是研究利用计量经济模型来分析多元过程 $\boldsymbol{r}_t$. 本章还讨论简化 $\boldsymbol{r}_t$ 的动态结构或降低 $\boldsymbol{r}_t$ 的维数的方法.

299

前几章中讨论的许多模型和方法可以直接推广为多元的情形. 但是有些情形下，这种推广需要注意一些问题. 在某些情形下，需要新的模型和方法来处理多个收益率之间的复杂关系. 我们还考虑了寻找影响不同资产收益率的公共因子的方法. 我们的讨论着重于直观性和实用性. 对多元时间序列分析的统计理论，读者可以参考 Lütkepohl(1991) 和 Reinsel(1993).

8.1 弱平稳与交叉相关矩阵

考虑一个 k 维时间序列 $\boldsymbol{r}_t=(r_{1t}, r_{2t}, \cdots, r_{kt})'$. 如果它的一阶矩与二阶矩是不随时间变化的，则称序列 $\boldsymbol{r}_t$ 是弱平稳的. 特别地，弱平稳序列 $\boldsymbol{r}_t$ 的均值向量和协方差矩阵不随时间变化. 除了明确地说明为相反的情况，我们都假定金融资产的收益序列是弱平稳的.

对一个弱平稳时间序列 $\boldsymbol{r}_t$，我们定义它的均值向量和协方差矩阵为

$$\boldsymbol{\mu}=E(\boldsymbol{r}_t), \quad \boldsymbol{\Gamma}_0=E[(\boldsymbol{r}_t-\boldsymbol{\mu}_t)(\boldsymbol{r}_t-\boldsymbol{\mu}_t)'] \tag{8.1}$$

这里的期望是对 r_t 的联合分布对每个分量取期望．均值 $\boldsymbol{\mu}$ 是由 r_t 的分量的无条件期望组成的 k 维向量．协方差矩阵 $\boldsymbol{\Gamma}_0$ 是 $k\times k$ 矩阵．$\boldsymbol{\Gamma}_0$ 的第 i 个对角线上的元素是 r_{it} 的方差，而 $\boldsymbol{\Gamma}_0$ 的第 (i, j) 个元素是 r_{it} 与 r_{jt} 的协方差．需要时，我们记为 $\boldsymbol{\mu}=(\mu_1, \cdots, \mu_k)'$，$\boldsymbol{\Gamma}_0=(\Gamma_{ij}(0))$．

8.1.1 交叉相关矩阵

令 $\boldsymbol{D}$ 表示由 $r_{it}(i=1, \cdots, k)$ 的标准差构成的 $k\times k$ 对角矩阵．换句话说，

$$\boldsymbol{D} = \operatorname{diag}\left\{\sqrt{\Gamma_{11}(0)}, \cdots, \sqrt{\Gamma_{kk}(0)}\right\}.$$

$\boldsymbol{r}_t$ 的同步或延迟为 0 的交叉相关矩阵定义为

$$\boldsymbol{\rho}_0 \equiv [\rho_{ij}(0)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_0\boldsymbol{D}^{-1},$$

更具体地，$\boldsymbol{\rho}_0$ 的第 (i, j) 个元素为

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\operatorname{Cov}(r_{it}, r_{jt})}{\operatorname{std}(r_{it})\operatorname{std}(r_{jt})},$$

它是 r_{it} 与 r_{jt} 间的相关系数．在时间序列分析中，这样一个相关系数称为一个同时或同步相关系数，因为它是两个序列在同一时刻 t 的相关性．很容易看出 $\rho_{ij}(0)=\rho_{ji}(0)$，$-1\leqslant\rho_{ij}(0)\leqslant1$；且 $\rho_{ii}(0)=1(1\leqslant i, j\leqslant k)$．这样，$\boldsymbol{\rho}(0)$ 是具有单位对角元素的对称矩阵． [300]

多元时间序列分析中一个重要的主题是分量序列之间的引导-延迟关系(lead-lag)．为此，用交叉相关矩阵来衡量时间序列之间线性依赖的强度．$\boldsymbol{r}_t$ 的延迟 l 的交叉协方差矩阵定义为

$$\boldsymbol{\Gamma}_l \equiv [\Gamma_{ij}(l)] = E[(\boldsymbol{r}_t-\boldsymbol{\mu})(\boldsymbol{r}_{t-l}-\boldsymbol{\mu})'] \tag{8.2}$$

其中 $\boldsymbol{\mu}$ 是 $\boldsymbol{r}_t$ 的均值向量．因此，$\boldsymbol{\Gamma}_l$ 的第 (i, j) 个元素为 r_{it} 与 $r_{j,t-l}$ 间的协方差．对弱平稳序列，交叉协方差矩阵 $\boldsymbol{\Gamma}_l$ 是 l 的函数，而不是时间指数 t 的函数．

$\boldsymbol{r}_t$ 的延迟 l 的交叉相关矩阵(CCM) 定义为

$$\boldsymbol{\rho}_l \equiv [\rho_{ij}(l)] = \boldsymbol{D}^{-1}\boldsymbol{\Gamma}_l\boldsymbol{D}^{-1} \tag{8.3}$$

同前面一样，$\boldsymbol{D}$ 是由单个序列 r_{it} 的标准差构成对角矩阵．由定义，

$$\rho_{ij}(l) = \frac{\Gamma_{ij}(l)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{\operatorname{Cov}(r_{it}, r_{j,t-l})}{\operatorname{std}(r_{it})\operatorname{std}(r_{jt})} \tag{8.4}$$

是 r_{it} 与 $r_{j,t-l}$ 的相关系数．当 $l>0$ 时，此相关系数衡量了 r_{it} 对发生在 t 时刻以前的 $r_{j,t-l}$ 的线性依赖．因此，如果 $\rho_{ij}(l)\neq0$，且 $l>0$，我们就说序列 r_{jt} 在延迟 l 处引导着序列 r_{it}．类似地，$\rho_{ji}(l)$ 衡量了 r_{jt} 对 $r_{i,t-l}$ 的线性依赖，并且如果 $\rho_{ji}(l)\neq0$ 且 $l>0$，我们就说 r_{it} 在延迟 l 处引导着序列 r_{jt}．方程(8.4)还表明对角元素 $\rho_{ii}(l)$ 仅仅为 r_{it} 的延迟 l 的自相关系数．

根据上述讨论，我们得到当 $l>0$ 时，交叉相关阵的一些重要性质．首先，一般地，对 $i\neq j$，有 $\rho_{ij}(l)\neq\rho_{ji}(l)$，因为这两个相关系数衡量的是 $\{r_{it}\}$ 与 $\{r_{jt}\}$ 之间的不同的线性关系．因此，$\boldsymbol{\Gamma}_l$ 与 $\boldsymbol{\rho}_l$ 一般是不对称的．第二，由 $\operatorname{Cov}(r_{it}, r_{j,t-l})=\operatorname{Cov}(r_{j,t-l}, r_{it})$，以及弱平稳性假定 $\operatorname{Cov}(r_{j,t-l}, r_{it})=\operatorname{Cov}(r_{jt}, r_{i,t+l})=\operatorname{Cov}[r_{jt}, r_{i,t-(-l)}]$，我们有 $\Gamma_{ij}(l)=\Gamma_{ji}(-l)$．因为 $\Gamma_{ji}(-l)$ 为矩阵 $\boldsymbol{\Gamma}_{-l}$ 的第 (j, i) 个元素，且这个等式对 $1\leqslant i, j\leqslant k$ 成立，所以我们有 $\boldsymbol{\Gamma}_l=\boldsymbol{\Gamma}'_{-l}$，$\boldsymbol{\rho}_l=\boldsymbol{\rho}'_{-l}$．因此，与一元情形不同，对一般的向量时间序列来说，当 $l>0$ 时，$\boldsymbol{\rho}_l\neq\boldsymbol{\rho}_{-l}$．因为 $\boldsymbol{\rho}_l=\boldsymbol{\rho}'_{-l}$，所以在实际中，只考虑 $l\geqslant0$ 时的交叉相关矩阵 $\boldsymbol{\rho}_l$ 就足够了． [301]

8.1.2 线性相依性

联合考虑一个弱平稳向量时间序列的交叉相关矩阵 $\{\boldsymbol{\rho}_l \mid l=0, 1, \cdots\}$ 包含下面的信息：

1)对角元素$\{\rho_{ii}(l) \mid l=0, 1, \cdots\}$是$r_{it}$的自相关函数.

2)非对角元素$\rho_{ij}(0)$衡量的是r_{it}与r_{jt}之间的同步线性关系.

3)对$l>0$，非对角元素$\rho_{ij}(l)$衡量的是r_{it}对过去值$r_{j,t-1}$的线性依赖.

因此，如果对所有的$l>0$，都有$\rho_{ij}(l)=0$，则r_{it}并不线性地依赖于r_{jt}序列的任何过去值$r_{j,t-l}$.

一般地，两个时间序列$\{r_{it}\}$与$\{r_{jt}\}$之间的线性关系可以概括如下：

1)如果对于所有的$l\geqslant 0$，都有$\rho_{ij}(l)=\rho_{ji}(l)=0$，则$r_{it}$与$r_{jt}$没有线性关系.

2)如果$\rho_{ij}(0)\neq 0$，则r_{it}与r_{jt}是同步相关的.

3)如果对于所有的$l>0$，$\rho_{ij}(l)=0$，且$\rho_{ji}(l)=0$，则r_{it}与r_{jt}没有引导-延迟关系. 这时称这两个序列是分离的.

4)如果对于所有的$l>0$，$\rho_{ij}(l)=0$，但是对某些$v>0$，有$\rho_{ji}(v)\neq 0$，则从r_{it}到r_{jt}有一个单向关系. 在这种情形，r_{it}并不依赖于r_{jt}的任何过去值，但是r_{jt}却依赖于r_{it}的某些过去值.

5)如果对某些$l>0$，$\rho_{ij}(l)\neq 0$；而且对某些$v>0$，$\rho_{ji}(v)\neq 0$，则r_{it}与r_{jt}之间具有一种反馈关系.

前面陈述的条件都是充分条件. 研究时间序列之间关系的更加有效的方法是对序列构造一个多元模型，因为一个恰当指定的模型同时考虑了时间序列间的序列相关性及交叉相关性.

8.1.3 样本交叉相关矩阵

给定数据$\{\boldsymbol{r}_t \mid t=1, \cdots, T\}$，其交叉协方差矩阵$\boldsymbol{\Gamma}_l$可以通过下式估计

$$\hat{\boldsymbol{\Gamma}}_l = \frac{1}{T}\sum_{t=l+1}^{T}(\boldsymbol{r}_t - \bar{\boldsymbol{r}})(\boldsymbol{r}_{t-l} - \bar{\boldsymbol{r}})', \qquad l \geqslant 0 \tag{8.5}$$

这里$\bar{\boldsymbol{r}} = \sum_{t=1}^{T}\boldsymbol{r}_t/T$为样本均值向量. 交叉相关矩阵$\boldsymbol{\rho}_l$估计为

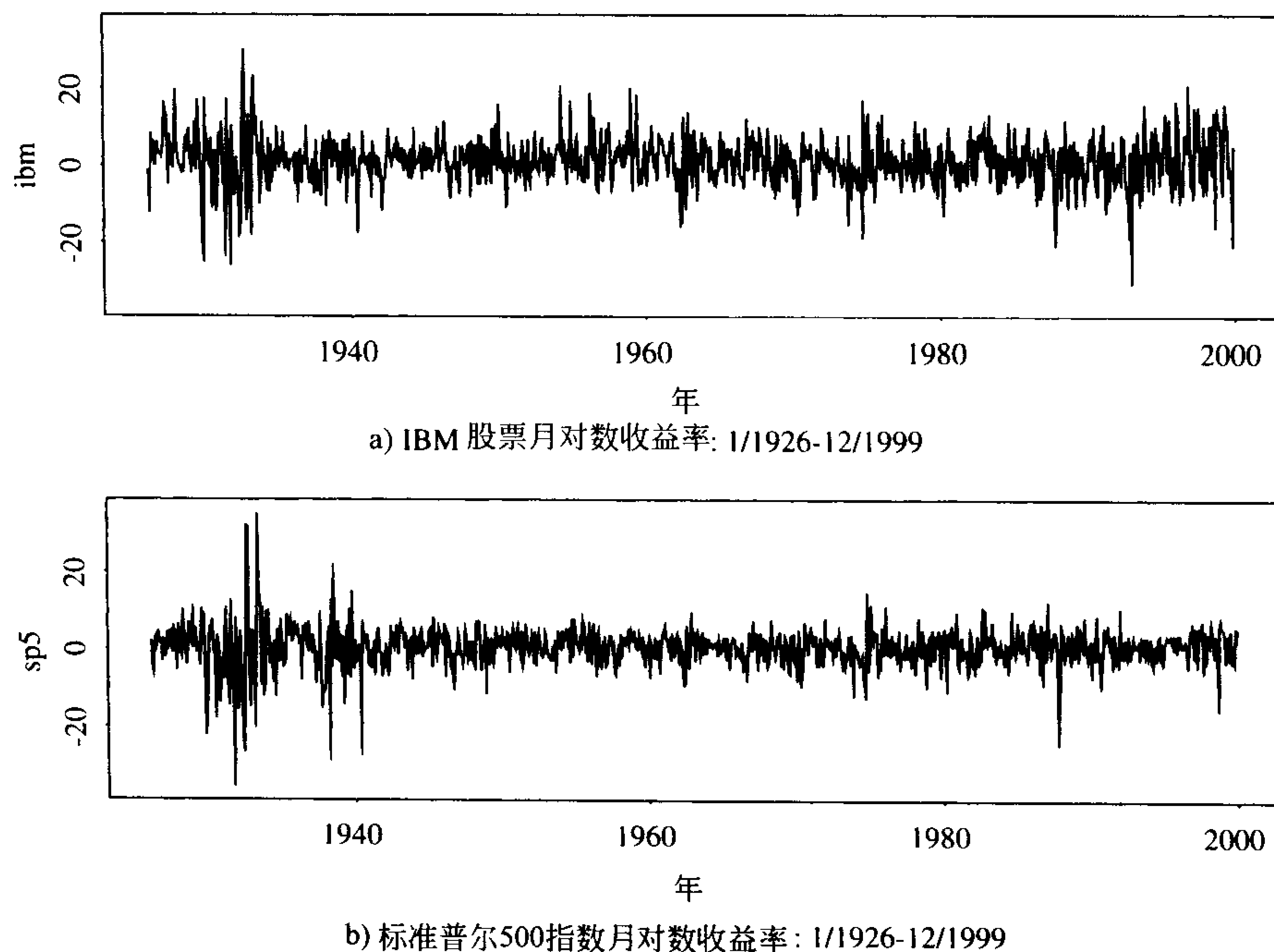
302

$$\hat{\boldsymbol{\rho}}_l = \hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\Gamma}}_l\hat{\boldsymbol{D}}^{-1}, \qquad l \geqslant 0 \tag{8.6}$$

其中$\hat{\boldsymbol{D}}$是分量序列的样本标准差构成的$k\times k$对角矩阵.

类似于一元情形，样本交叉相关矩阵$\hat{\boldsymbol{\rho}}_l$在各种假定之下的渐近性质都已经被研究了；例如可以参见 Fuller(1976，第 6 章). 这个估计是相合的，但是对于有限样本是有偏的. 对于资产收益率序列，$\hat{\boldsymbol{\rho}}_l$的有限样本分布相当复杂，部分原因是由于条件异方差与高峰度的出现. 如果需要交叉相关的有限样本分布，我们建议利用适当的自助(bootstrap)重新抽样方法得到分布的渐近估计. 对于许多应用而言，$\hat{\rho}_{ij}(l)$方差的一个粗糙估计就足够了.

例 8.1 考虑 IBM 股票与标准普尔 500 指数从 1926 年 1 月至 1999 年 12 月的月对数收益率，共 888 个观测值. 这个收益率包括了红利支付，并且是以百分比表示的. 分别用r_{1t}和r_{2t}表示 IBM 股票与标准普尔 500 指数的收益率. 这两个收益率组成了一个二元的时间序列$\boldsymbol{r}_t=(r_{1t}, r_{2t})'$. 图 8-1 给出了同样尺度的$\boldsymbol{r}_t$的时间图，图 8-2 给出了两个序列的一些散点图. 这些图表明这两个收益率序列是相关的. 事实上，两个收益率之间的样本同步相关系数是 0.64，这在 5%水平下是统计显著的. 然而，1 步延迟的交叉相关若有的话也是很弱的.



a) IBM股票月对数收益率: 1/1926-12/1999

b) 标准普尔500指数月对数收益率: 1/1926-12/1999

图 8-1 IBM 股票与标准普尔 500 指数从 1926 年 1 月至 1999 年 12 月的月对数收益率的时间图

303

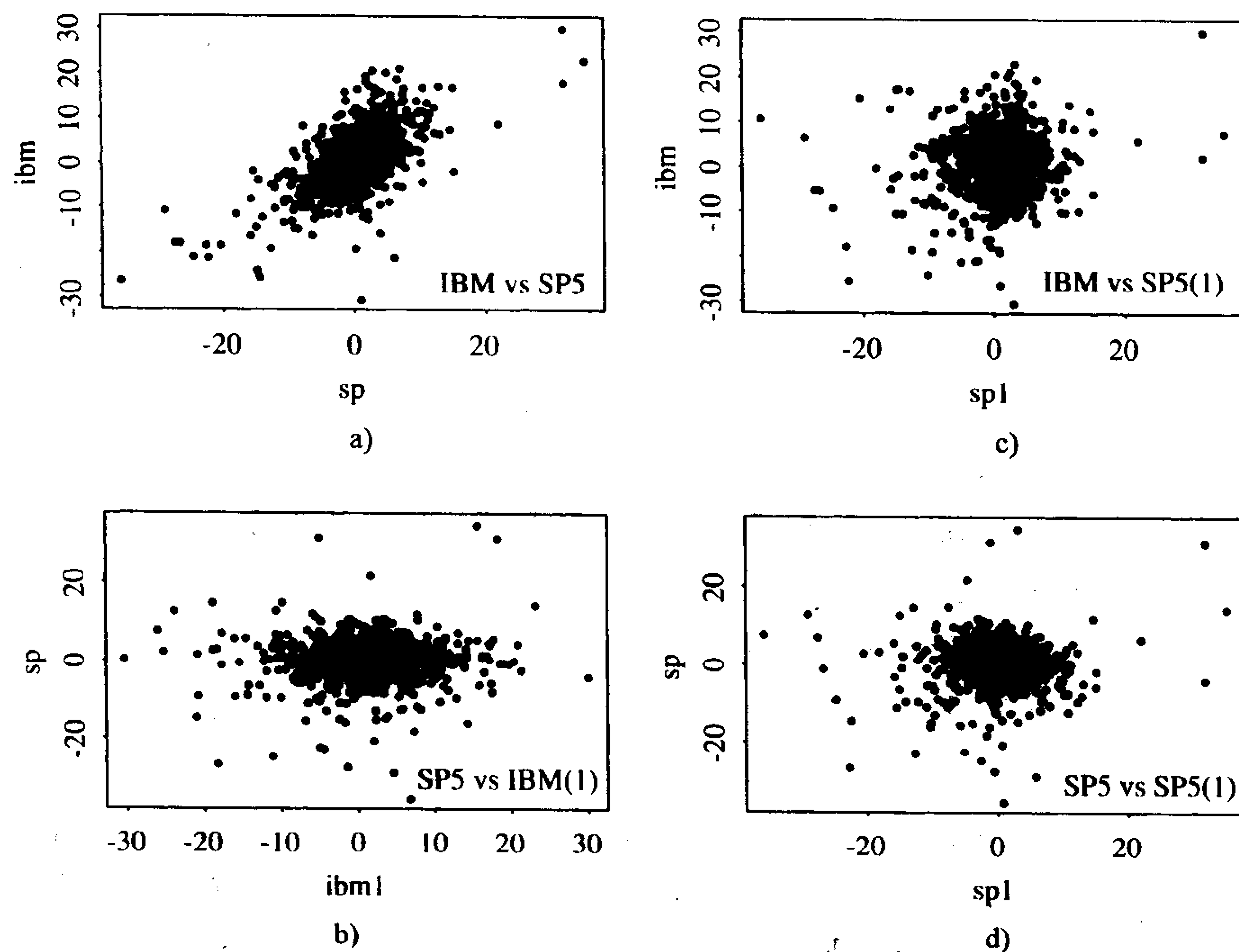


图 8-2 IBM 股票与标准普尔 500 指数的月对数收益率的一些散点图：a)同步图；b)标准普尔 500 对 1 步延迟 IBM；c)IBM 对 1 步延迟标准普尔 500；d)标准普尔 500 对 1 步延迟标准普尔 500

表 8-1 给出了两个序列的一些概括性统计量与交叉相关矩阵. 对一个二元序列，每个 CCM 都是一个 2×2 矩阵，包含四个相关系数. 实证经验表明很难同时吸收许多的交叉相关矩阵，尤其是当维数 k 大于 3 时. 为了克服这个困难，我们利用 Tiao 和 Box(1981) 的简化记号，并且定义一个简单的交叉相关矩阵，它包含三个符号"＋"，"－"，和"."，其中

1)"＋"意味着相应的相关系数大于或等于 $2/\sqrt{T}$，

2)"－"意味着相应的相关系数小于或等于 $-2/\sqrt{T}$，

3)"."意味着相应的相关系数介于 $-2/\sqrt{T}$ 与 $2/\sqrt{T}$ 之间.

304 这里 $2/\sqrt{T}$ 是假定 $\boldsymbol{r}_t$ 为一个白噪声时，在 5%渐近水平下样本相关系数的临界值.

表 8-1

(a)概括统计量

最小变动价位	均值	标准误差	偏度	超出峰度	最小值	最大值
IBM	1.240	6.729	−0.237	1.917	−30.37	30.10
SP5	0.537	5.645	−0.521	8.117	−35.58	35.22

(b)交叉相关矩阵

延迟 1	延迟 2	延迟 3	延迟 4	延迟 5
0.08　0.10	0.02　−0.06	−0.02　−0.07	−0.02　−0.03	0.00　0.07
0.04　0.08	0.02　−0.02	−0.07　−0.11	0.04　0.02	0.00　0.08

(c)简单记号

延迟 1	延迟 2	延迟 3	延迟 4	延迟 5
$\begin{bmatrix} + & + \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ - & - \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & + \end{bmatrix}$

注：IBM 股票和标准普尔 500 指数的月对数收益率的概括性统计量和交叉相关矩阵. 时间段是 1926 年 1 月至 1999 年 12 月.

表 8-1(c)给出了 IBM 股票与标准普尔 500 指数的月对数收益率的简化 CCM. 很容易看出在 5%渐近水平上显著的交叉相关主要出现在延迟 3 和延迟 5 处. 检查这两个延迟的样本 CCM 可见：(a)标准普尔 500 指数收益率在延迟 1 与 3 有某些边际自相关性；(b)IBM 股票收益率对标准普尔 500 指数的过去收益率的线性依赖关系很弱. 后一观测是根据交叉相关在延迟 1 与延迟 3 的 CCM 的第(1，2)个元素的显著性.

图 8-3 显示了两个序列的样本自相关与交叉相关. 因为对延迟为 0 的 ACF 是对称的，只给出具有正的延迟值的 ACF. 因为使用标准普尔 500 指数收益率的延迟值来计算交叉相关，所以图 8-3c 中与正延迟相联系的图给出了 IBM 股票收益率对过去的标准普尔 500 指数收益率的依赖性；与负延迟相联系的图给出了标准普尔 500 指数收益率对过去的 IBM 股票收益率的依赖性. 图中的水平线是样本自相关与交叉相关系数的渐近两个标准差的上下限. 从图中可以看出，两个收益率序列之间的动态关系是非常弱的，但是它们的同步相关却是统计显著的.

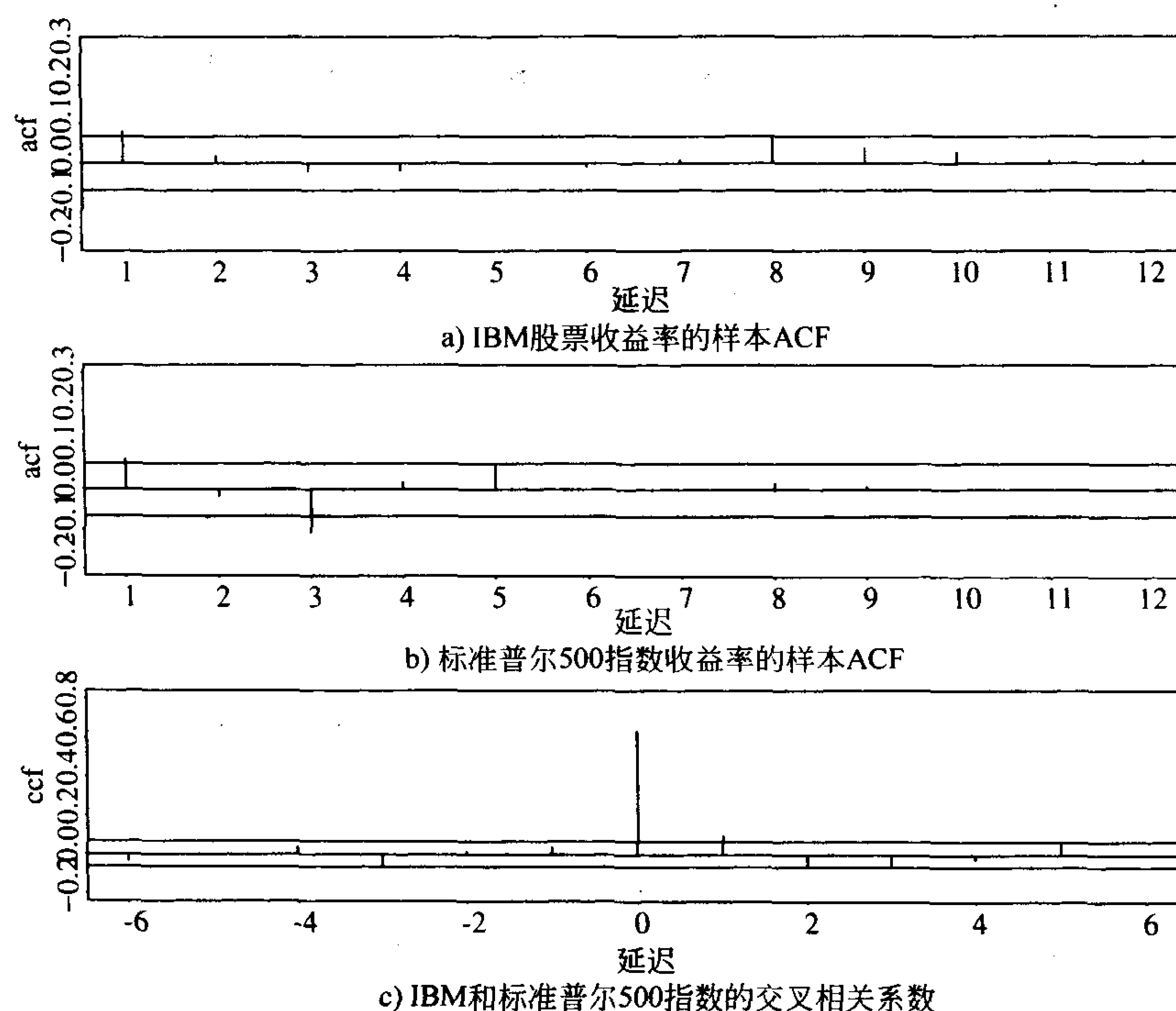


图 8-3　两个月对数收益率的样本自相关和交叉相关函数

例 8.2　考虑期限为 30 年、20 年、10 年、5 年和 1 年的美国政府债券指数的月简单收益率指数. 数据来自于 CRSP 数据库，样本期间为 1942 年 1 月至 1999 年 12 月，共 696 个观测值. 令 $\boldsymbol{r}_t=(r_{1t},\ \cdots,\ r_{5t})'$为具有递减的剩余期限的收益率序列. 图 8-4 给出了同样尺寸的 $\boldsymbol{r}_t$ 的时间图. 1 年债券收益率的可变性远远小于较长期限的收益率的可变性. 数据的样本均值和标准差分别为

$$\hat{\boldsymbol{\mu}}=10^{-2}(0.43,0.45,0.45,0.46,0.44)'$$
$$\hat{\boldsymbol{\sigma}}=10^{-2}(2.53,2.43,1.97,1.39,0.53)'$$

序列的同步相关矩阵为

$$\hat{\boldsymbol{\rho}}_0=\begin{bmatrix}1.00 & 0.98 & 0.92 & 0.85 & 0.63\\ 0.98 & 1.00 & 0.91 & 0.86 & 0.64\\ 0.92 & 0.91 & 1.00 & 0.90 & 0.68\\ 0.85 & 0.86 & 0.90 & 1.00 & 0.82\\ 0.63 & 0.64 & 0.68 & 0.82 & 1.00\end{bmatrix}$$

出现下列现象并不奇怪：(a)序列有高度的同步相关性. (b)长期债券之间的相关要高于短期债券之间的相关.

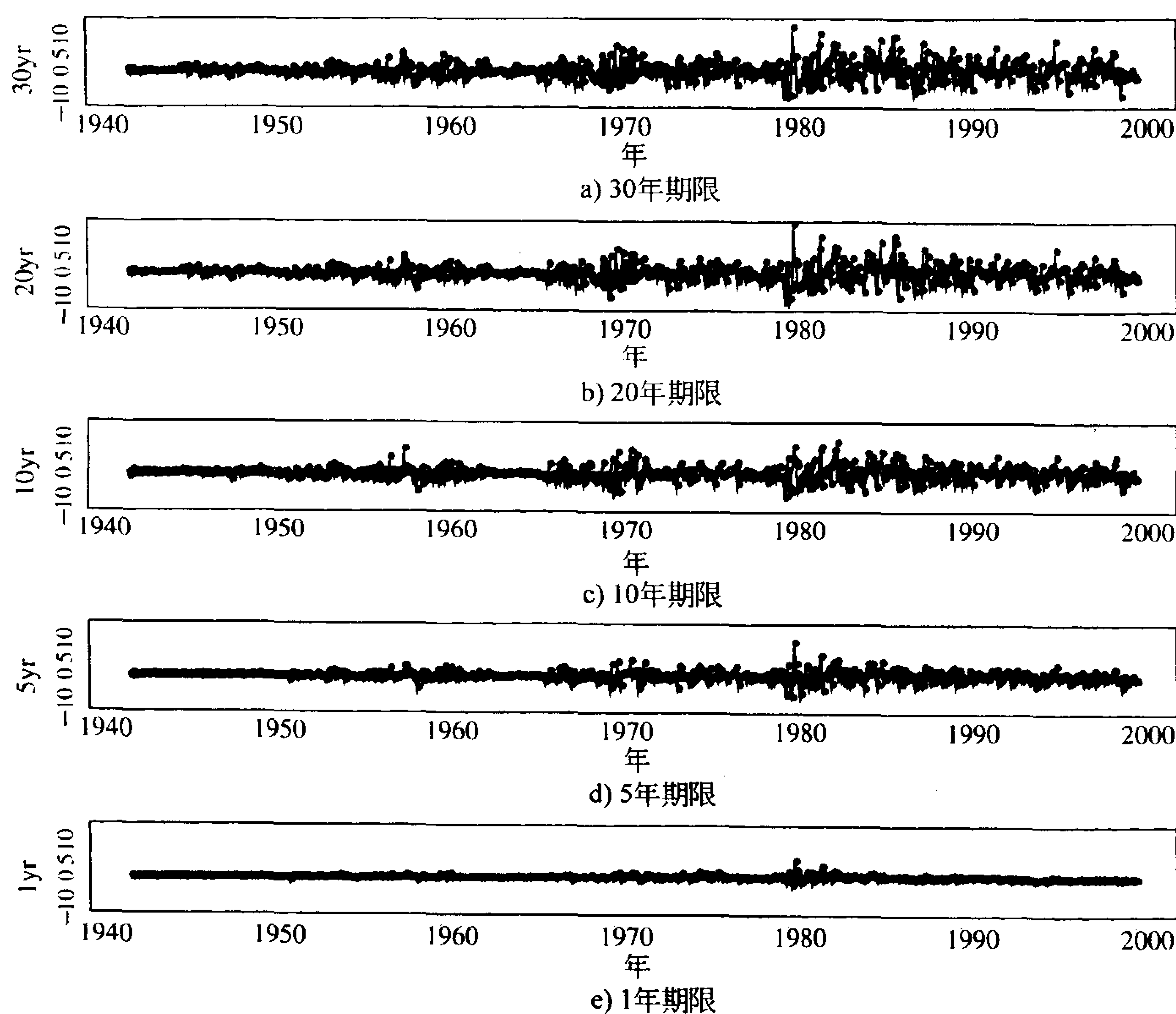


图 8-4 美国政府债券 5 个指数的月简单收益率的时间图，期限分别是 30 年、20 年、10 年、5 年和 1 年．样本时间段是 1942 年 1 月至 1999 年 12 月

表 8-2 给出了 $\boldsymbol{r}_t$ 的延迟 1 与延迟 2 的交叉相关矩阵以及相应的简化矩阵．绝大多数显著的交叉相关在延迟 1 处，而且 5 个收益序列看上去是交互相关的．另外，1 年期债券收益率的延迟 1 与延迟 2 的样本 ACF 高于其他的较长期限序列的 ACF.

表 8-2 美国政府债券 5 个指数的月简单收益率的样本交叉相关矩阵．数据周期从 1942 年 1 月至 1999 年 12 月

延迟 1					延迟 2				
(a)交叉相关									
0.10	0.08	0.11	0.12	0.16	−0.01	0.00	0.00	−0.03	0.03
0.10	0.08	0.12	0.14	0.17	−0.01	0.00	0.00	−0.04	0.02
0.09	0.08	0.09	0.13	0.18	0.01	0.01	0.01	−0.02	0.07
0.14	0.12	0.15	0.14	0.22	−0.02	−0.01	0.00	−0.04	0.07
0.17	0.15	0.21	0.22	0.40	−0.02	0.00	0.02	0.02	0.22

（续）

延迟 1	延迟 2
(b)简化交叉相关矩阵	
$\begin{bmatrix} + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & + \end{bmatrix}$

305 ~ 307

8.1.4 多元混成检验

Hosking(1980，1981) 与 Li 和 McLeod(1981) 已经把一元的 Ljung-Box 统计量$Q(m)$推广到多元情形. 对一个多元序列，检验统计量的原假设为 H_0：$\boldsymbol{\rho}_1=\cdots=\boldsymbol{\rho}_m=\mathbf{0}$，备择假设为 H_a：对某些 $i\in\{1,\cdots,m\}$，$\boldsymbol{\rho}_i\neq\mathbf{0}$. 这样，利用这个统计量来检验向量序列 $\boldsymbol{r}_t$ 没有自相关和交叉相关性. 假定检验统计量的形式为

$$Q_k(m)=T^2\sum_{l=1}^{m}\frac{1}{T-l}\mathrm{tr}(\hat{\boldsymbol{\Gamma}}_l'\hat{\boldsymbol{\Gamma}}_0^{-1}\hat{\boldsymbol{\Gamma}}_l\hat{\boldsymbol{\Gamma}}_0^{-1}) \tag{8.7}$$

其中 T 为样本容量，k 为 $\boldsymbol{r}_t$ 的维数，$tr(\mathbf{A})$是矩阵 $\mathbf{A}$ 的迹，即 $\mathbf{A}$ 的对角线元素的和. 在原假设以及一些正则条件下，$Q_k(m)$渐近服从一个自由度为 k^2m 的 χ^2 分布.

注释：$Q_k(m)$统计量可以根据样本交叉相关矩阵$\hat{\boldsymbol{\rho}}_l$ 改写，但是这种表示涉及到克罗内克(Kronecker)积$\otimes$与本章附录 A 讨论的矩阵的拉直. 利用这些算子，我们有

$$Q_k(m)=T^2\sum_{l=1}^{m}\frac{1}{T-l}\boldsymbol{b}_l'(\hat{\boldsymbol{\rho}}_0^{-1}\otimes\hat{\boldsymbol{\rho}}_0^{-1})\boldsymbol{b}_l.$$

这里 $\boldsymbol{b}_l=\mathrm{vec}(\hat{\boldsymbol{\rho}}_l')$. 更具体地，Li 和 McLeod(1981) 提出的检验统计量为

$$Q_k^*(m)=T\sum_{l=1}^{m}\boldsymbol{b}_l'(\hat{\boldsymbol{\rho}}_0^{-1}\otimes\hat{\boldsymbol{\rho}}_0^{-1})\boldsymbol{b}_l+\frac{k^2m(m+1)}{2T},$$

它渐近等价于 $Q_k(m)$.

对例 8.1 中的 IBM 股票与标准普尔 500 指数的二元月对数收益率应用$Q_k(m)$统计量，我们有 $Q_2(1)=9.81$，$Q_2(5)=47.06$，$Q_2(10)=71.65$. 根据自由度为 4、20 和 40 的渐近 χ^2 分布，这些 $Q_2(m)$统计量的 p 值都接近于 0. 因此，混成检验确认了二元收益率序列中前后依赖的存在. 对于例 8.2 中债券指数的 5 维月简单收益率，我们有 $Q_5(5)=1\,065.63$，与自由度为 125 的 χ^2 分布相比较，是高度显著的.

$Q_k(m)$统计量是对 $\boldsymbol{r}_t$ 的前 m 个交叉相关矩阵的一个联合检验. 如果原假设被拒绝，那么我们必须对序列建立一个多元模型来研究分量序列之间的引导-延迟关系. 下面我们讨论一些简单的向量模型，他们在给多元金融时间序列的线性动态结构建模时很有用.

308

8.2 向量自回归模型

在给资产收益率建模时，一个简单有用的向量模型是向量自回归（VAR)模型. 多元时间序列 $\boldsymbol{r}_t$ 是一个一阶的 VAR 过程，或者简称为 VAR(1)，如果它服从下面的模型

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{a}_t \tag{8.8}$$

这里 $\boldsymbol{\phi}_0$ 是一个 k 维向量，$\boldsymbol{\Phi}$ 是一个 $k\times k$ 矩阵，$\{\boldsymbol{a}_t\}$是序列不相关的随机向量序列，其均值为 0，协方差矩阵为Σ. 实际应用中，要求协方差矩阵Σ是正定的；否则，可以简化 $\boldsymbol{r}_t$ 的维数. 文献中，通常假定 $\boldsymbol{a}_t$ 是多元正态的.

考虑二元情形(即 $k=2$，$\boldsymbol{r}_t=(r_{1t}, r_{2t})'$ 且 $\boldsymbol{a}_t=(a_{1t}, a_{2t})'$). VAR(1)模型包含下面两个方程：

$$r_{1t} = \phi_{10} + \Phi_{11}r_{1,t-1} + \Phi_{12}r_{2,t-1} + a_{1t}$$
$$r_{2t} = \phi_{20} + \Phi_{21}r_{1,t-1} + \Phi_{22}r_{2,t-1} + a_{2t},$$

其中 Φ_{ij} 是 $\boldsymbol{\Phi}$ 的第(i, j)个元素，ϕ_{i0} 是 $\boldsymbol{\phi}_0$ 的第 i 个元素. 根据第一个方程，Φ_{12} 表示的是在 $r_{1,t-1}$存在时，r_{1t}对 $r_{2,t-1}$的线性依赖，因此 Φ_{12} 为给定 $r_{1,t-1}$时，$r_{2,t-1}$对 r_{1t}的条件效应. 如果 $\Phi_{12}=0$，那么 r_{1t}并不依赖于 $r_{2,t-1}$，而且模型表明 r_{1t}只依赖于自己的过去值. 类似地，如果 $\Phi_{21}=0$，那么第二个方程表明给定 $r_{2,t-1}$时，r_{2t}并不依赖于 $r_{1,t-1}$.

联合考虑这两个方程. 如果 $\Phi_{12}=0$，但是 $\Phi_{21}\neq 0$，那么从 r_{1t}到 r_{2t}有一个单向关系. 如果 $\Phi_{12}=\Phi_{21}=0$，那么 r_{1t}与 r_{2t}是分离的. 如果 $\Phi_{12}\neq 0$，且 $\Phi_{21}\neq 0$，那么这两个序列之间有一个反馈关系.

一般地，系数矩阵 $\boldsymbol{\Phi}$ 度量 $\boldsymbol{r}_t$ 的动态相依性. r_{1t}与 r_{2t}之间的同步关系可以通过 $\boldsymbol{a}_t$ 的协方差矩阵Σ的非对角线元素 σ_{12}来反映. 如果 $\sigma_{12}=0$，那么这两个分量序列之间没有同步线性关系. 在计量经济文献中，方程 (8.8) 中的 VAR(1) 模型称为简化形式的模型，因为它没有清楚地给出分量序列之间的同步相依性. 如果有必要，我们可以通过对简化形式的模型作一个简单的线性变换得到包含同步关系的一个显式表达. 因为Σ是正定的，所以存在一个对角线上元素全为 1 的下三角矩阵 $\boldsymbol{L}$ 以及对角矩阵 $\boldsymbol{G}$，满足 $\Sigma=\boldsymbol{LGL}'$；见附录 A 中的楚列斯基(Cholesky)分解. 因此，$\boldsymbol{L}^{-1}\Sigma(\boldsymbol{L}')^{-1}=\boldsymbol{G}$.

定义 $\boldsymbol{b}_t=(b_{1t}, \cdots, b_{kt})'=\boldsymbol{L}^{-1}\boldsymbol{a}_t$，则

309

$$E(\boldsymbol{b}_t)=\boldsymbol{L}^{-1}E(\boldsymbol{a}_t)=\boldsymbol{0},\quad \mathrm{Cov}(\boldsymbol{b}_t)=\boldsymbol{L}^{-1}\Sigma(\boldsymbol{L}^{-1})'=\boldsymbol{L}^{-1}\Sigma(\boldsymbol{L}')^{-1}=\boldsymbol{G}.$$

因为 $\boldsymbol{G}$ 是对角矩阵，所以 $\boldsymbol{b}_t$ 的分量是不相关的. 在方程(8.8)中的两端同时左乘 $\boldsymbol{L}^{-1}$，我们得到

$$\boldsymbol{L}^{-1}\boldsymbol{r}_t = \boldsymbol{L}^{-1}\boldsymbol{\phi}_0 + \boldsymbol{L}^{-1}\boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{L}^{-1}\boldsymbol{a}_t = \boldsymbol{\phi}_0^* + \boldsymbol{\Phi}^*\boldsymbol{r}_{t-1} + \boldsymbol{b}_t \tag{8.9}$$

其中 $\boldsymbol{\phi}_0^*=\boldsymbol{L}^{-1}\boldsymbol{\phi}_0$ 是一个 k 维向量，而且 $\boldsymbol{\Phi}^*=\boldsymbol{L}^{-1}\boldsymbol{\Phi}$ 是 $k\times k$ 矩阵. 因为这个特殊的矩阵结构，$\boldsymbol{L}^{-1}$的第 k 行具有形式(w_{k1}, w_{k2}, $\cdots$, $w_{k,k-1}$, 1). 因此，模型 (8.9) 的第 k 个方程是

$$r_{kt}+\sum_{i=1}^{k-1}w_{ki}r_{it}=\phi_{k,0}^{*}+\sum_{i=1}^{k}\Phi_{ki}^{*}r_{i,t-1}+b_{kt} \tag{8.10}$$

这里$\phi_{k,0}^{*}$是ϕ_0^{*} 的第 k 个元素，Φ_{ki}^{*}是$\boldsymbol{\Phi}^{*}$的第(k，i)个元素．因为对$1\leqslant i<k$，b_{kt}与b_{it}是不相关的，所以方程(8.10)明确给出了r_{kt}对r_{it}的同步线性依赖性，这里$1\leqslant i<k-1$．在计量经济文献中，该方程称为r_{kt}的一个结构方程．

对$\boldsymbol{r}_t$的任何其他分量r_{it}，我们可以对 VAR(1)模型进行重排，使得r_{it}变为$\boldsymbol{r}_t$的最后一个分量．可以利用前面的变换方法得到r_{it}的一个结构方程．因此，简化形式的模型(8.8)等价于计量经济文献中使用的结构形式．在时间序列分析中，通常使用简化形式的模型，原因有两个：第一个原因是易于估计，第二个，也是主要的原因，是在预测时不能用同步相关性．

例 8.3 为了说明从简化形式模型到结构方程的变换，考虑二元 AR(1)模型

$$\begin{bmatrix}r_{1t}\\r_{2t}\end{bmatrix}=\begin{bmatrix}0.2\\0.4\end{bmatrix}+\begin{bmatrix}0.2&0.3\\-0.6&1.1\end{bmatrix}\begin{bmatrix}r_{1,t-1}\\r_{2,t-1}\end{bmatrix}+\begin{bmatrix}a_{1t}\\a_{2t}\end{bmatrix}\quad \Sigma=\begin{bmatrix}2&1\\1&1\end{bmatrix}$$

对这个特殊的自协方差矩阵Σ，下三角矩阵

$$\boldsymbol{L}^{-1}=\begin{bmatrix}1.0&0.0\\-0.5&1.0\end{bmatrix}$$

给出了一个楚列斯基分解(即$\boldsymbol{L}^{-1}\Sigma(\boldsymbol{L}')^{-1}$是一个对角阵)．将前面的二元 AR(1)模型左乘上因子$\boldsymbol{L}^{-1}$，我们得到

$$\begin{bmatrix}1.0&0.0\\-0.5&1.0\end{bmatrix}\begin{bmatrix}r_{1t}\\r_{2t}\end{bmatrix}=\begin{bmatrix}0.2\\0.3\end{bmatrix}+\begin{bmatrix}0.2&0.3\\-0.7&0.95\end{bmatrix}\begin{bmatrix}r_{1,t-1}\\r_{2,t-1}\end{bmatrix}+\begin{bmatrix}b_{1t}\\b_{2t}\end{bmatrix}$$

$$\boldsymbol{G}=\begin{bmatrix}2&0\\0&0.5\end{bmatrix}$$

310

这里$\boldsymbol{G}=\mathrm{Cov}(\boldsymbol{b}_t)$．这个变换模型的第二个方程为

$$r_{2t}=0.3+0.5r_{1t}-0.7r_{1,t-1}+0.95r_{2,t-1}+b_{2t}.$$

明确给出了r_{2t}对r_{1t}的线性依赖．

重新安排$\boldsymbol{r}_t$中元素的顺序，则二元 AR(1)模型变为

$$\begin{bmatrix}r_{2t}\\r_{1t}\end{bmatrix}=\begin{bmatrix}0.4\\0.2\end{bmatrix}+\begin{bmatrix}1.1&-0.6\\0.3&0.2\end{bmatrix}\begin{bmatrix}r_{2,t-1}\\r_{1,t-1}\end{bmatrix}+\begin{bmatrix}a_{2t}\\a_{1t}\end{bmatrix},\quad \Sigma=\begin{bmatrix}1&1\\1&2\end{bmatrix}$$

在Σ的楚列斯基分解中需要的下三角矩阵变为

$$\boldsymbol{L}^{-1}=\begin{bmatrix}1.0&0.0\\-1.0&1.0\end{bmatrix}$$

将前面重排的 VAR(1)模型前乘以$\boldsymbol{L}^{-1}$，我们得到

$$\begin{bmatrix}1.0&0.0\\-1.0&1.0\end{bmatrix}\begin{bmatrix}r_{2t}\\r_{1t}\end{bmatrix}=\begin{bmatrix}0.4\\-0.2\end{bmatrix}+\begin{bmatrix}1.1&-0.6\\-0.8&0.8\end{bmatrix}\begin{bmatrix}r_{2,t-1}\\r_{1,t-1}\end{bmatrix}+\begin{bmatrix}c_{1t}\\c_{2t}\end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

其中 $\boldsymbol{G}=\text{Cov}(\boldsymbol{c}_t)$. 现在第二个方程给出

$$r_{1t} = -0.2 + 1.0r_{2t} + 0.8r_{2,t-1} + 0.8r_{1,t-1} + c_{2t}$$

这个方程又一次明确给出了 r_{1t}对 r_{2t}的同步线性依赖性.

8.2.1 VAR(1)模型的平稳性条件和矩

假定方程(8.8)中的 VAR(1)模型是弱平稳的. 对这个模型取期望，利用$E(\boldsymbol{a}_t)=\boldsymbol{0}$，我们得到

$$E(\boldsymbol{r}_t) = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}E(\boldsymbol{r}_{t-1})$$

因为 $E(\boldsymbol{r}_t)$是不随时间变的，假定矩阵 $\boldsymbol{I}-\boldsymbol{\Phi}$ 是非奇异的，则我们有

$$\boldsymbol{\mu} \equiv E(\boldsymbol{r}_t) = (\boldsymbol{I}-\boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_0$$

其中 $\boldsymbol{I}$ 是 $k\times k$ 单位矩阵.

利用 $\boldsymbol{\phi}_0=(\boldsymbol{I}-\boldsymbol{\Phi})\boldsymbol{\mu}$，则方程(8.8)中的 VAR(1)模型可以写为

$$(\boldsymbol{r}_t - \boldsymbol{\mu}) = \boldsymbol{\Phi}(\boldsymbol{r}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{a}_t$$

令$\tilde{\boldsymbol{r}}_t=\boldsymbol{r}_t-\boldsymbol{\mu}$ 是均值修正的时间序列，则 VAR(1)模型变为

311

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}\tilde{\boldsymbol{r}}_{t-1} + \boldsymbol{a}_t \tag{8.11}$$

这个模型可以用来推导出 VAR(1)模型的性质. 通过重复代入，我们可以将方程(8.11)改写为

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{a}_t + \boldsymbol{\Phi}\boldsymbol{a}_{t-1} + \boldsymbol{\Phi}^2\boldsymbol{a}_{t-2} + \boldsymbol{\Phi}^3\boldsymbol{a}_{t-3} + \cdots$$

这个表示给出了 VAR(1)过程的几个特征：第一，因为 $\boldsymbol{a}_t$ 是前后无关的，从而 $\text{Cov}(\boldsymbol{a}_t, \boldsymbol{r}_{t-1})=\boldsymbol{0}$. 事实上，对所有的 $l>0$，$\boldsymbol{a}_t$ 与 $\boldsymbol{r}_{t-l}$都是不相关的. 由于这个原因，将 $\boldsymbol{a}_t$ 称为序列在时刻 t 时的一个抖动或新息. 可以证明，同 1 元的情形类似，对所有的时间序列模型，$\boldsymbol{a}_t$ 与过去值 $\boldsymbol{r}_{t-j}(j>0)$是不相关的. 第二，将这个表示右乘以 $\boldsymbol{a}_t'$后取期望，并利用过程 $\boldsymbol{a}_t$ 的序列不相关性，我们得到 $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t)=\boldsymbol{\Sigma}$. 第三，对一个 VAR(1)模型，$\boldsymbol{r}_t$ 以系数矩阵 $\boldsymbol{\Phi}^j$依赖于过去的新息 $\boldsymbol{a}_{t-j}$. 为了这种相依性有意义，当 $j\to\infty$时，$\boldsymbol{\Phi}^j$必须收敛到 0. 这意味着 $\boldsymbol{\Phi}$ 的 k 个特征值的模必须都小于 1；否则，当 $j\to\infty$时，$\boldsymbol{\Phi}^j$要么发散，要么会收敛到一个不为 0 的矩阵，事实上，如果假定 $\boldsymbol{a}_t$ 的协方差矩阵存在，那么要求 $\boldsymbol{\Phi}$ 的所有特征值的模都小于 1，就是 $\boldsymbol{r}_t$ 的弱平稳性的充分必要条件. 注意到对 1 元 AR(1)情形，这个平稳性条件简化为$|\phi|<1$. 第四，利用这个表示，我们有

$$\text{Cov}(\boldsymbol{r}_t)=\boldsymbol{\Gamma}_0=\boldsymbol{\Sigma}+\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}'+\boldsymbol{\Phi}^2\boldsymbol{\Sigma}(\boldsymbol{\Phi}^2)'+\cdots=\sum_{i=0}^{\infty}\boldsymbol{\Phi}^i\boldsymbol{\Sigma}(\boldsymbol{\Phi}^i)'$$

其中 $\boldsymbol{\Phi}^0=\boldsymbol{I}$ 是 $k\times k$ 单位阵.

将方程(8.11)两端右乘以$\tilde{\boldsymbol{r}}_{t-l}'$后取期望，并利用对 $j>0$，有 $\text{Cov}(\boldsymbol{a}_t, \boldsymbol{r}_{t-j})=E(\boldsymbol{a}_t\tilde{\boldsymbol{r}}_{t-j}')=\boldsymbol{0}$ 的结果，我们得到

$$E(\tilde{\boldsymbol{r}}_t\tilde{\boldsymbol{r}}_{t-l})' = \boldsymbol{\Phi}E(\tilde{\boldsymbol{r}}_{t-1}\tilde{\boldsymbol{r}}_{t-l})', \qquad l>0$$

因此，

$$\boldsymbol{\Gamma}_l = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{l-1}, \qquad l > 0 \tag{8.12}$$

其中 $\boldsymbol{\Gamma}_j$ 是 $\boldsymbol{r}_t$ 的延迟 j 的交叉协方差矩阵. 这个结果又是 1 元 AR(1)过程的一个推广. 通过重复代入，方程(8.12)表明：$\boldsymbol{\Gamma}_l=\boldsymbol{\Phi}^l\boldsymbol{\Gamma}_0$，$l>0$.

8.2.2 向量 AR(p)模型

VAR(1)模型可以直接推广到 VAR(p)模型. 称时间序列 $\boldsymbol{r}_t$ 服从一个 VAR(p)模型，如果它满足

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{r}_{t-p} + \boldsymbol{a}_t, \qquad p > 0, \tag{8.13}$$

312

其中 $\boldsymbol{\phi}_0$ 与 $\boldsymbol{a}_t$ 如前所定义，$\boldsymbol{\Phi}_j$ 是 $k\times k$ 矩阵. 利用向后推移算子 B，VAR(p) 模型可以写为

$$(\boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t$$

其中 $\boldsymbol{I}$ 是 $k\times k$ 单位矩阵. 这个表示可以写为一个更紧凑的形式

$$\boldsymbol{\Phi}(B)\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{a}_t$$

其中 $\boldsymbol{\Phi}(B)=\boldsymbol{I}-\boldsymbol{\Phi}_1B-\cdots-\boldsymbol{\Phi}_pB^p$ 是一个矩阵多项式. 如果 $\boldsymbol{r}_t$ 是弱平稳的，并假定逆存在，则我们有

$$\boldsymbol{\mu} = E(\boldsymbol{r}_t) = (\boldsymbol{I} - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_p)^{-1}\boldsymbol{\phi}_0 = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\phi}_0$$

令$\tilde{\boldsymbol{r}}_t=\boldsymbol{r}_t-\boldsymbol{\mu}$，则 VAR($p$)模型变为

$$\tilde{\boldsymbol{r}}_t = \boldsymbol{\Phi}_1\tilde{\boldsymbol{r}}_{t-1} + \cdots + \boldsymbol{\Phi}_p\tilde{\boldsymbol{r}}_{t-p} + \boldsymbol{a}_t \tag{8.14}$$

利用这个方程和与对 VAR(1)模型同样的方法，我们得到

- $\text{Cov}(\boldsymbol{r}_t, \boldsymbol{a}_t)=\Sigma$，$\Sigma$是 $\boldsymbol{a}_t$ 的协方差矩阵；
- $\text{Cov}(\boldsymbol{r}_{t-l}, \boldsymbol{a}_t)=\boldsymbol{0}$，对 $l>0$；
- $\boldsymbol{\Gamma}_l=\boldsymbol{\Phi}_1\boldsymbol{\Gamma}_{l-1}+\cdots+\boldsymbol{\Phi}_p\boldsymbol{\Gamma}_{l-p}$，对 $l>0$.

最后一个性质称为 VAR(p) 模型的矩方程，它是 1 元 AR(p)模型的 Yule-Walker 方程的多元形式.

理解方程 (8.13)式的这个 VAR(p)模型性质的一个简单方法是利用方程(8.8)中 VAR(1)模型的结果，可以通过将 $\boldsymbol{r}_t$ 的 VAR(p)模型变换为一个 kp 维的 VAR(1)模型得到. 具体地，令 $\boldsymbol{x}_t=(\tilde{\boldsymbol{r}}'_{t-p+1}, \tilde{\boldsymbol{r}}'_{t-p+2}, \cdots, \tilde{\boldsymbol{r}}'_t)'$，$\boldsymbol{b}_t=(0, \cdots, 0, \boldsymbol{a}'_t)'$为两个 kp 维的过程，$\boldsymbol{b}_t$ 的均值是 0，协方差矩阵是一个 $kp\times kp$ 的矩阵，只有右下角的元素不为 0，用Σ表示. 这样，对于 $\boldsymbol{r}_t$ 的 VAR(p)模型可写成

$$\boldsymbol{x}_t = \boldsymbol{\Phi}^*\boldsymbol{x}_{t-1} + \boldsymbol{b}_t \tag{8.15}$$

其中 $\boldsymbol{\Phi}^*$ 是一个 $kp\times kp$ 矩阵：

$$\boldsymbol{\Phi}^* = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} \\ \boldsymbol{\Phi}_p & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_{p-2} & \boldsymbol{\Phi}_{p-3} & \cdots & \boldsymbol{\Phi}_1 \end{bmatrix},$$

313

这里 $\mathbf{0}$ 与 $\boldsymbol{I}$ 分别是 $k\times k$ 的零矩阵和单位矩阵. 文献中称 $\boldsymbol{\Phi}^*$ 为矩阵多项式 $\boldsymbol{\Phi}(B)$ 的伴随矩阵(companion matrix).

方程(8.15)是 $\boldsymbol{x}_t$ 的一个 VAR(1)模型，它包含了 $\boldsymbol{r}_t$ 作为它的最后 k 个分量. 现在可以通过方程(8.15)用上一小节中显示的 VAR(1)模型的结果来推导 VAR(p)模型的性质. 例如，由定义，$\boldsymbol{x}_t$ 是弱平稳的当且仅当 $\boldsymbol{r}_t$ 是弱平稳的. 因此，方程 (8.13)中 VAR(p)模型弱平稳的充要条件是方程(8.15)中 $\boldsymbol{\Phi}^*$ 的所有特征值的模都小于 1.

在金融时间序列分析中特别关心的是 VAR(p)模型的系数矩阵 $\boldsymbol{\Phi}_l$ 的结构. 例如，如果对所有的 l，$\boldsymbol{\Phi}_l$ 的第(i，j)个元素 $\Phi_{ij}(l)$都是 0，则 r_{it}不依赖于 r_{jt}的过去值. 这样，系数矩阵 $\boldsymbol{\Phi}_l$ 的结构提供了 $\boldsymbol{r}_t$ 的分量之间的引导-延迟关系的信息.

8.2.3 建立一个 VAR(p)模型

我们继续利用阶的识别、估计以及模型检验的迭代程序来对一个给定的时间序列构造一个向量 AR 模型. 1 元序列 PACF(偏自相关函数)的概念可以推广到识别向量序列的阶 p. 考虑下面的相邻的 VAR 模型：

$$
\begin{aligned}
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \boldsymbol{a}_t \\
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{r}_{t-2} + \boldsymbol{a}_t \\
\vdots &= \vdots \\
\boldsymbol{r}_t &= \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1\boldsymbol{r}_{t-1} + \cdots + \boldsymbol{\Phi}_i\boldsymbol{r}_{t-i} + \boldsymbol{a}_t \\
\vdots &= \vdots
\end{aligned}
\tag{8.16}
$$

这些模型的参数可以通过普通最小二乘方法(OLS)来估计. 这在多元统计分析中称为多元线性回归估计，见 Johnson 和 Wichern(1998).

对方程组 (8.16) 的第 i 个方程，令$\hat{\boldsymbol{\Phi}}_j^{(i)}$ 表示$\boldsymbol{\Phi}_j$ 的 OLS 估计，$\hat{\boldsymbol{\phi}}_0^{(i)}$ 表示 $\boldsymbol{\phi}_0$ 的估计，这里上标(i)用来表示估计是针对 VAR(i)模型的，则残差为

$$
\hat{\boldsymbol{a}}_t^{(i)} = \boldsymbol{r}_t - \hat{\boldsymbol{\phi}}_0^{(i)} - \hat{\boldsymbol{\Phi}}_1^{(i)}\boldsymbol{r}_{t-1} - \cdots - \hat{\boldsymbol{\Phi}}_i^{(i)}\boldsymbol{r}_{t-i}
$$

对 $i=0$，残差定义为$\hat{\boldsymbol{r}}_t^{(0)}=\boldsymbol{r}_t-\bar{\boldsymbol{r}}$，这里$\bar{\boldsymbol{r}}$为 $\boldsymbol{r}_t$ 的样本均值. 残差协方差矩阵定义为

314

$$
\hat{\boldsymbol{\Sigma}}_i = \frac{1}{T-2i-1}\sum_{t=i+1}^{T}\hat{\boldsymbol{a}}_t^{(i)}(\hat{\boldsymbol{a}}_t^{(i)})' \quad i\geqslant 0
\tag{8.17}
$$

为了确定阶 p，可以对 $l=1$，2，…检验假设 H_0：$\boldsymbol{\Phi}_l=\mathbf{0}$，对备择假设 H_a：$\boldsymbol{\Phi}_l\neq\mathbf{0}$. 例如，利用方程组(8.16)中第一个方程，我们可以检验假设 H_0：$\boldsymbol{\Phi}_1=\mathbf{0}$ 对备择假设 H_a：$\boldsymbol{\Phi}_1\neq\mathbf{0}$. 检验统计量为：

$$
M(1) = -\left(T-k-2\frac{1}{2}\right)\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}_1|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)
$$

其中$\hat{\boldsymbol{\Sigma}}_i$ 定义见方程(8.17)，$|\boldsymbol{A}|$表示矩阵$\boldsymbol{A}$ 的行列式. 在一些正则条件下，检验统计量 $M(1)$ 渐近服从自由度为 k^2 的 χ^2 分布；见 Tiao 和 Box(1981).

一般地，我们利用方程组(8.16)中的第 i 个和第 $i-1$ 个方程来检验 H_0：$\boldsymbol{\Phi}_i=\mathbf{0}$ 对 H_a：$\boldsymbol{\Phi}_i\neq\mathbf{0}$，即检验一个 VAR($i$)模型对一个 VAR($i-1$)模型. 检验统计量为

$$M(i)=-\left(T-k-i-\frac{3}{2}\right)\ln\left(\frac{|\hat{\Sigma}_i|}{|\hat{\Sigma}_{i-1}|}\right) \tag{8.18}$$

$M(i)$渐近服从自由度为 k^2 的 χ^2 分布.

另一种选择，可以利用 AIC(Akaike 信息准则)或它的变形来选择阶 p. 假定 $\boldsymbol{a}_t$ 是多元正态的，并考虑方程组 (8.16) 中的第 i 个方程，可以用最大似然(ML)方法来估计这个模型. 对于 AR 模型，OLS 估计$\hat{\phi}_0$ 和$\hat{\boldsymbol{\Phi}}_j$ 等价于(条件)ML 估计. 然而，Σ的估计之间有所区别. Σ的 ML 估计为

$$\tilde{\Sigma}_i=\frac{1}{T}\sum_{t=i+1}^{T}\hat{\boldsymbol{a}}_t^{(i)}[\hat{\boldsymbol{a}}_t^{(i)}]' \tag{8.19}$$

VAR(i)模型在正态假定下的 AIC 定义为 $AIC(i)=\ln(|\tilde{\Sigma}_i|)+\frac{2k^2 i}{T}$. 对一个给定的向量时间序列，可以选择 AR 阶 p，使它满足 $AIC(p)=\min_{1\leqslant i\leqslant p_0}AIC(i)$，其中 p_0 是一个预先指定的正整数.

例 8.4 假定例 8.1 中讨论的 IBM 股票和标准普尔 500 指数的月对数收益率的 2 元序列服从一个 VAR 模型，我们对数据利用 $M(i)$统计量和 AIC. 表 8-3 给出了这些统计量的值. 两个统计量都表明对数据运用 VAR(3)模型可能是合适的. 在 5%水平下，$M(i)$统计量在延迟 1、3 和 5 处都是边际显著的. AIC 的最小值在阶 3 处出现. 对这个特例，$M(i)$统计量在 1%水平下是不显著的，确认了前面的观测：两个收益率序列之间的动态线性相依性是弱的. 315

表 8-3

Order	1	2	3	4	5	6
$M(i)$	9.81	8.93	12.57	6.08	9.56	2.80
AIC	6.757	6.756	6.750	6.753	6.751	6.756

注：IBM 股票和标准普尔 500 指数从 1926 年 1 月到 1999 年 12 月的月对数收益率的定阶统计量. 自由度为 4 的 χ^2 分布的 5%和 1%临界值分别为 9.5 和 13.3.

估计和模型检验

对一个指定的 VAR 模型，可以利用普通最小二乘法或最大似然方法来估计参数. 这两个方法是渐近等价的. 在一些正则条件下，估计是渐近正态的；见 Reinsel (1993). 应该对一个拟合模型的任何可能的合适性进行仔细的检验. 可以对残差序列利用 $Q_k(m)$统计量来检验残差之间没有序列相关或交叉相关的假定. 对一个拟合的 VAR(p)模型，残差的 $Q_k(m)$统计量渐近服从自由度为 k^2m-g 的 χ^2 分布，这里 g 为 AR 系数矩阵中待估参数的数量.

例 8.4(续) 表 8-4(a)给出了对 IBM 股票和标准普尔 500 指数的月对数收益率的 2 元序列的 VAR(3)模型的估计结果，指定的模型具有形式

$$\boldsymbol{r}_t=\boldsymbol{\phi}_0+\boldsymbol{\Phi}_1\boldsymbol{r}_{t-1}+\boldsymbol{\Phi}_3\boldsymbol{r}_{t-3}+\boldsymbol{a}_t \tag{8.20}$$

其中 r_t 的第一分量表示 IBM 股票的收益率. 对这个特例，我们仅仅使用了延迟 1 和 3 的 AR 系数矩阵，这是因为数据之间的弱序列相关性. 一般地，当 $M(i)$统计量和 AIC 准则指定一个 VAR(3)模型时，应该使用所有的 3 个 AR 延迟. 表 8-4(b)给出了将一些统计上不显著的参数设为 0 以后的估计结果. 表 8-4(b)中对拟合模型的残差序列的 $Q_k(m)$统计量给出 $Q_2(4)=18.17$，$Q_2(8)=41.26$. 因为拟合的 VAR(3)模型在 AR 系数矩阵中具有四个参数，所以这两个 $Q_k(m)$统计量分别渐近服从自由度为 12 和 28 的 χ^2 分布. 检验统计量的 p 值分别为 0.111 和 0.051，因此在 5%的显著性水平下，模型是合理的. 如同 1 元分析中显示的，收益率序列很可能具有条件异方差性. 我们在第 9 章中将讨论多元波动率.

表 8-4　IBM 股票和标准普尔 500 指数从 1926 年 1 月到 1999 年 12 月的月对数收益率的VAR(3)的估计结果

参数	ϕ_0	$\boldsymbol{\Phi}_1$		$\boldsymbol{\Phi}_3$		Σ	
			(a)整个模型				
估计值	1.20	0.011	0.108	0.039	−0.112	44.44	23.51
	0.58	−0.013	0.084	−0.007	−0.105	23.51	31.29
标准差	0.23	0.043	0.051	0.044	0.052		
	0.19	0.036	0.043	0.037	0.044		
			(b)简化的模型				
估计值	1.24	0	0.117	0	−0.083	44.48	23.51
	0.57	0	0.073	0	−0.109	23.51	31.29
标准差	0.23	—	0.040	—	0.040		
	0.19	—	0.033	—	0.033		

从表 8-4(b)中拟合的模型，我们观测到：

(a)两个新息序列之间的同步相关系数为 $23.51/\sqrt{44.48\times 31.29}=0.63$，如所料想，它很接近于 r_{1t}与 r_{2t}之间的样本相关系数.

(b)两个对数收益率序列具有正的显著的均值，蕴涵着两个序列的对数价格在该数据范围内具有上升趋势.

(c)这个模型表明

$$\mathrm{IBM}_t = 1.24 + 0.117\mathrm{SP5}_{t-1} - 0.083\mathrm{SP5}_{t-3} + a_{1t},$$

$$\mathrm{SP5}_t = 0.57 + 0.073\mathrm{SP5}_{t-1} - 0.109\mathrm{SP5}_{t-3} + a_{2t}.$$

因此在 5%显著水平下，从标准普尔 500 指数月收益率到 IBM 月收益率有一个单向动态关系. 如果标准普尔 500 指数代表美国股票市场，则 IBM 收益率受该市场过去运动的影响. 然而，即使两个收益率有实质上的同步相关性，但是 IBM 股票收益率的过去运动并不显著地影响美国股票市场. 最后，拟合的模型可以写为：

$$\begin{bmatrix}\mathrm{IBM}_t\\ \mathrm{SP5}_t\end{bmatrix}=\begin{bmatrix}1.24\\ 0.57\end{bmatrix}+\begin{bmatrix}0.117\\ 0.073\end{bmatrix}\mathrm{SP5}_{t-1}-\begin{bmatrix}0.083\\ 0.109\end{bmatrix}\mathrm{SP5}_{t-3}+\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}$$

这表示 $\mathrm{SP5}_t$ 是该 2 元序列的驱动因子(driven factor).

预测

将合理建立的模型看作是真实的模型，则可以应用1元分析中同样的方法来产生预测与相应预测误差的标准差. 对一个VAR(p)模型，在时间起始点h的1步向前预测为$\boldsymbol{r}_h(1)=\phi_0+\sum_{i=1}^{p}\boldsymbol{\Phi}_i\boldsymbol{r}_{h+1-i}$，相应的预测误差为$\boldsymbol{e}_h(1)=\boldsymbol{a}_{h+1}$. 预测误差的协方差矩阵为$\Sigma$. 如果$\boldsymbol{r}_t$是弱平稳的，则当预测期$l$增加时，$l$步向前预测$\boldsymbol{r}_h(l)$收敛到它的均值向量$\boldsymbol{\mu}$.

316～317

总之，建立一个VAR模型涉及三个步骤：(a)利用检验统计量$M(i)$或Akaike信息准则来识别阶；(b)利用最小二乘法估计指定的模型(某些情况下，可以通过消除统计上不显著的参数来重新估计这个模型)；(c)利用残差的$Q_k(m)$统计量来检验拟合模型的充分性. 残差序列的其他特征，例如条件异方差和异常值，也可以检验，如果拟合的模型是合适的，则可以用它来进行预测.

8.3 向量滑动平均模型

一个阶为q的向量滑动平均模型或VMA(q)具有形式

$$\boldsymbol{r}_t=\boldsymbol{\theta}_0+\boldsymbol{a}_t-\boldsymbol{\Theta}_1\boldsymbol{a}_{t-1}-\cdots-\boldsymbol{\Theta}_q\boldsymbol{a}_{t-q}\quad\text{或}\quad\boldsymbol{r}_t=\boldsymbol{\theta}_0+\boldsymbol{\Theta}(B)\boldsymbol{a}_t \tag{8.21}$$

这里$\boldsymbol{\theta}_0$是k维向量，$\boldsymbol{\Theta}_i$是$k\times k$个矩阵，$\boldsymbol{\Theta}(B)=\boldsymbol{I}-\boldsymbol{\Theta}_1B-\cdots-\boldsymbol{\Theta}_qB^q$是向后推移算子$B$的MA矩阵多项式. 类似于1元的情形，假定$\boldsymbol{a}_t$的协方差矩阵$\Sigma$存在，则VMA($q$)过程是弱平稳的. 将方程(8.21)取期望，得到$\boldsymbol{\mu}=E(\boldsymbol{r}_t)=\boldsymbol{\theta}_0$. 这样，对VMA模型而言，常数向量$\boldsymbol{\theta}_0$是$\boldsymbol{r}_t$的均值向量.

令$\tilde{\boldsymbol{r}}_t=\boldsymbol{r}_t-\boldsymbol{\theta}_0$为均值修正的VAR($q$)过程. 那么利用方程(8.21)以及$\{\boldsymbol{a}_t\}$是序列不相关的事实，我们有：

1)$\text{Cov}(\boldsymbol{r}_t,\ \boldsymbol{a}_t)=\Sigma$，

2)$\boldsymbol{\Gamma}_0=\Sigma+\boldsymbol{\Theta}_1\Sigma\boldsymbol{\Theta}_1'+\cdots+\boldsymbol{\Theta}_q\Sigma\boldsymbol{\Theta}_q'$，

3)若$l>q$，则$\boldsymbol{\Gamma}_l=0$，

4)若$1\leqslant l\leqslant q$，则$\boldsymbol{\Gamma}_l=\sum_{j=l}^{q}\boldsymbol{\Theta}_j\Sigma\boldsymbol{\Theta}_{j-l}'$，这里$\boldsymbol{\Theta}_0=-\boldsymbol{I}$.

因为对$l>q$，有$\boldsymbol{\Gamma}_l=\boldsymbol{0}$，所以VMA($q$)过程$\boldsymbol{r}_t$的交叉相关矩阵(CCM)满足

$$\boldsymbol{\rho}_l=\boldsymbol{0},\qquad l>q. \tag{8.22}$$

因此，类似于1元情形，可以利用样本CCM来识别一个VMA过程的阶.

为了更好地理解VMA过程，让我们考虑2元MA(1)模型.

$$\boldsymbol{r}_t=\boldsymbol{\theta}_0+\boldsymbol{a}_t-\boldsymbol{\Theta}\boldsymbol{a}_{t-1}=\boldsymbol{\mu}+\boldsymbol{a}_t-\boldsymbol{\Theta}\boldsymbol{a}_{t-1} \tag{8.23}$$

这里为了简便，消除了$\boldsymbol{\Theta}_1$的下标. 这个模型可以清楚地写为：

318

$$\begin{bmatrix}r_{1t}\\ r_{2t}\end{bmatrix}=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}+\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}-\begin{bmatrix}\Theta_{11} & \Theta_{12}\\ \Theta_{21} & \Theta_{22}\end{bmatrix}\begin{bmatrix}a_{1,t-1}\\ a_{2,t-1}\end{bmatrix} \tag{8.24}$$

它说明当前的收益率序列 $\boldsymbol{r}_t$ 仅仅依赖于当前的和过去的抖动. 因此，这个模型是一个有限记忆模型.

考虑方程(8.24)中 r_{1t}所满足的方程，参数 Θ_{12}表示 $a_{1,t-1}$存在时，r_{1t}对 $a_{2,t-1}$的线性依赖. 如果 $\Theta_{12}=0$，则 r_{1t}并不依赖于 a_{2t}的延迟值，因此也不依赖于 r_{2t}的延迟值. 类似地，如果 $\Theta_{21}=0$，那么 r_{2t}并不依赖于 r_{1t}的过去值. $\boldsymbol{\Theta}$ 的非对角元素反映了分量序列之间的动态依赖性. 对这个简单的 VMA(1)模型，我们可以将 r_{1t}与 r_{2t}之间的关系分类如下：

1)如果 $\Theta_{12}=\Theta_{21}=0$，则它们是分离的序列.

2)如果 $\Theta_{12}=0$，但 $\Theta_{21}\neq0$，则从 r_{1t}到 r_{2t}，有一个单向动态关系. 如果 $\Theta_{21}=0$，但 $\Theta_{12}\neq 0$，则相反的单向动态关系成立.

3)如果 $\Theta_{12}\neq0$，且 $\Theta_{21}\neq0$，则 r_{1t}与 r_{2t}之间有一个反馈关系.

最后，r_{it}之间的同步关系与 a_{it}之间的同步关系是一样的. 前面的分类可以推广到 VMA(q)模型.

估计

与 VAR 模型不同，VMA 模型的估计更加复杂；见 Hillmer 和 Tiao(1979) Lütkepohl (1991)及其参考文献. 对似然方法，有两种可以利用. 第一个方法是条件似然方法，它假定对 $t\leqslant0$，有 $\boldsymbol{a}_t=\boldsymbol{0}$. 第二个方法是精确似然方法，它将 $t\leqslant0$ 时的 $\boldsymbol{a}_t$ 看作模型的一个附加参数，为了获得估计问题的一些洞察，我们考虑方程(8.23)中的 VMA(1)模型. 假定数据为$\{\boldsymbol{r}_t \mid t=1, \cdots, T\}$，且 $\boldsymbol{a}_t$ 是多元正态的. 对一个 VMA(1)模型，这个数据依赖于 $\boldsymbol{a}_0$.

条件 MLE

条件似然方法假定 $\boldsymbol{a}_0=\boldsymbol{0}$. 在这个假定下，将模型改写为 $\boldsymbol{a}_t=\boldsymbol{r}_t-\boldsymbol{\theta}_0+\boldsymbol{\Theta}\boldsymbol{a}_{t-1}$，我们可以递推地计算抖动 $\boldsymbol{a}_t$，有

$$\boldsymbol{a}_1=\boldsymbol{r}_1-\boldsymbol{\theta}_0,\quad \boldsymbol{a}_2=\boldsymbol{r}_2-\boldsymbol{\theta}_0+\boldsymbol{\Theta}_1\boldsymbol{a}_1,\quad \cdots$$

因此，数据的对数似然函数变为：

$$f(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_T \mid \boldsymbol{\theta}_0, \boldsymbol{\Theta}_1, \Sigma)=\prod_{t=1}^{T}\frac{1}{(2\pi)^{k/2}\mid\Sigma\mid^{1/2}}\exp\left[-\frac{1}{2}\boldsymbol{a}_t'\Sigma^{-1}\boldsymbol{a}_t\right]$$

[319] 可以估计它来得到参数估计.

精确 MLE

对精确似然方法，$\boldsymbol{a}_0$ 是一个未知向量，为了估计似然函数，必须根据数据进行估计，为了简便，令$\tilde{\boldsymbol{r}}_t=\boldsymbol{r}_t-\boldsymbol{\theta}_0$ 为均值修正序列. 利用$\tilde{\boldsymbol{r}}_t$ 与方程(8.23)，我们有

$$\boldsymbol{a}_t=\tilde{\boldsymbol{r}}_t+\boldsymbol{\Theta}\boldsymbol{a}_{t-1} \tag{8.25}$$

通过重复代入，$\boldsymbol{a}_0$ 与所有的$\tilde{\boldsymbol{r}}_t$ 都相关，因为

$$\begin{aligned}\boldsymbol{a}_1&=\tilde{\boldsymbol{r}}_1+\boldsymbol{\Theta}\boldsymbol{a}_0\\ \boldsymbol{a}_2&=\tilde{\boldsymbol{r}}_2+\boldsymbol{\Theta}\boldsymbol{a}_1=\tilde{\boldsymbol{r}}_2+\boldsymbol{\Theta}\tilde{\boldsymbol{r}}_1+\boldsymbol{\Theta}^2\boldsymbol{a}_0\\ \vdots&=\vdots\end{aligned}$$

$$\boldsymbol{a}_T = \tilde{\boldsymbol{r}}_T + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{T-1} + \cdots + \boldsymbol{\Theta}^{T-1}\tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}^T\boldsymbol{a}_0 \tag{8.26}$$

这样，如果给定 $\boldsymbol{\Theta}$ 和 $\boldsymbol{\theta}_0$，则 $\boldsymbol{a}_0$ 是数据的线性函数．这个结果使得我们可以利用数据与 $\boldsymbol{\theta}_0$ 以及 $\boldsymbol{\Theta}$ 的初始估计值来估计 $\boldsymbol{a}_0$，更具体地，给定 $\boldsymbol{\theta}_0$，$\boldsymbol{\Theta}$ 与数据，我们可以定义：

$$\boldsymbol{r}_t^* = \tilde{\boldsymbol{r}}_t + \boldsymbol{\Theta}\tilde{\boldsymbol{r}}_{t-1} + \cdots + \boldsymbol{\Theta}^{t-1}\tilde{\boldsymbol{r}}_1, \quad 对 \quad t = 1,2,\cdots,T$$

则方程组(8.26)可以改写为：

$$\begin{aligned}
\boldsymbol{r}_1^* &= -\boldsymbol{\Theta}\boldsymbol{a}_0 + \boldsymbol{a}_1 \\
\boldsymbol{r}_2^* &= -\boldsymbol{\Theta}^2\boldsymbol{a}_0 + \boldsymbol{a}_2 \\
\vdots &= \vdots \\
\boldsymbol{r}_T^* &= -\boldsymbol{\Theta}^T\boldsymbol{a}_0 + \boldsymbol{a}_T
\end{aligned}$$

这是多元线性回归的形式，参数向量为 $\boldsymbol{a}_0$，尽管 $\boldsymbol{a}_t$ 的协方差矩阵可能不是一个对角阵．如果 Σ 的初始估计值也可以得到，则可以用 Σ 的平方根矩阵 $\Sigma^{-1/2}$ 左乘上面系统中的每个方程．结果中的系统真的是一个多元线性回归，且可以利用 OLS 方法来得到 $\boldsymbol{a}_0$ 的一个估计，表示为 $\hat{\boldsymbol{a}}_0$．

利用估计 $\hat{\boldsymbol{a}}_0$，可以递推地计算扰动 $\boldsymbol{a}_t$，因为

$$\boldsymbol{a}_1 = \boldsymbol{r}_1 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\hat{\boldsymbol{a}}_0, \quad \boldsymbol{a}_2 = \boldsymbol{r}_2 - \boldsymbol{\theta}_0 + \boldsymbol{\Theta}\boldsymbol{a}_1, \cdots$$

另外，我们也可以从 $\{\boldsymbol{a}_t \mid t=0, \cdots, T\}$ 的联合分布中推出数据的精确似然函数．可以利用估计结果中的似然函数来得到精确的 ML 估计．

总之，精确似然方法的操作如下：给定 $\boldsymbol{\theta}_0$，$\boldsymbol{\Theta}$ 与 Σ 的初始估计，利用方程组(8.26)导出 $\boldsymbol{a}_0$ 的估计．利用方程(8.25)，这个估计又反过来可以用来递推地计算 $\boldsymbol{a}_t$，初始值为 $\boldsymbol{a}_1 = \tilde{\boldsymbol{r}}_1 + \boldsymbol{\Theta}\hat{\boldsymbol{a}}_0$，然后利用结果中的 $\{\boldsymbol{a}_t\}_{t=1}^T$ 估计数据的精确似然函数，以便更新 $\boldsymbol{\theta}_0$，$\boldsymbol{\Theta}$ 和 Σ 的估计．重复整个过程直至估计收敛．评估精确似然函数的这个迭代方法对一般的 VMA(q)模型也是适用的． 320

由前面的讨论知，精确似然方法比条件似然方法要求更多精深的计算．但是它提供了更精确的参数估计，尤其是当 $\boldsymbol{\Theta}$ 的某些特征值的模接近于 1 时．Hillmer 和 Tiao(1979)提供了 VMA 模型的条件似然估计与精确似然估计之间的一些比较．在多元时间序列分析中，如果怀疑数据可能过度差分，则精确最大似然方法变得很重要．过度差分可能在很多情况下发生(例如，协整系统中单个分量的差分；见后面对协整的讨论)．

总之，建立一个 VMA 模型需要三个步骤：

(a)利用样本交叉相关矩阵识别阶 q：对 VMA(q)模型，当 $l>q$ 时，$\boldsymbol{\rho}_l=\boldsymbol{0}$；

(b)利用条件似然方法或精确似然方法来估计指定的模型：当样本量不太大时，更喜欢用精确方法；

(c)应该检验所拟合模型的合适性(例如对残差序列利用 $Q_k(m)$ 统计量)．

最后，可以利用与 1 元 MA 模型同样的程序得到 VMA 模型的预测．

例 8.5 再次考虑 IBM 股票和标准普尔 500 指数从 1926 年 1 月到 1999 年 12 月，以百分

比表示的月对数收益率的二元序列. 因为显著的交叉相关主要在延迟 1 和延迟 3 发生，所以我们对数据采用这个 VMA(3)模型：

$$\boldsymbol{r}_t = \boldsymbol{\theta}_0 + \boldsymbol{a}_t - \boldsymbol{\Theta}_1 \boldsymbol{a}_{t-1} - \boldsymbol{\Theta}_3 \boldsymbol{a}_{t-3} \tag{8.27}$$

表 8-5 给出了这个模型的估计结果. 这个简化模型的残差的 $Q_k(m)$统计量给出 $Q_2(4)=17.25$ 且 $Q_2(8)=39.30$. 与自由度为 12 和 28 的 χ^2 分布比较，这些统计量的 p 值分别是 0.140 4 和 0.076 2. 这样，此模型在 5%的显著水平下是充分的.

由表 8-5，我们作出以下观测：

(1)对这个特例，条件似然估计与精确似然估计的差别很小，这并不奇怪. 因为样本量并不小，而且更重要的是数据的动态结构很弱.

(2)VMA(3)模型对序列提供了与例 8.4 的 VAR(3)模型同样的动态关系. IBM 股票的月对数收益率依赖标准普尔 500 指数过去的收益率. 相反地，市场收益率并不依赖于 IBM 股票的延迟收益率，换句话说，数据的动态结构是由市场收益率而并不是 IBM 收益率驱动的，然而，两个收益率序列间的同步相关性仍很强.

321

表 8-5

参数	θ_0	Θ_1		Θ_3		Σ	
(a)整个模型用条件似然法							
估计值	1.24	−0.013	−0.121	−0.038	0.108	44.48	23.52
	0.54	0.020	−0.101	0.014	0.105	23.52	31.20
标准差	0.24	0.043	0.051	0.044	0.052		
	0.18	0.036	0.043	0.036	0.043		
(b)整个模型用精确似然法							
估计值	1.24	−0.013	−0.121	−0.038	0.108	44.48	23.52
	0.54	0.020	−0.101	0.013	0.105	23.52	31.20
标准差	0.24	0.043	0.051	0.044	0.052		
	0.18	0.036	0.043	0.036	0.043		
(c)简化的模型用精确似然法							
估计值	1.24	0.000	−0.126	0.000	0.082	44.54	23.51
	0.54	0.000	−0.084	0.000	0.114	23.51	31.21
标准差	0.23	—	0.040	—	0.040		
	0.18	—	0.033	—	0.033		

注：IBM 股票和标准普尔 500 指数的月对数收益率的向量滑动平均模型(8.27)的估计结果. 时间区间是从 1926 年 1 月到 1999 年 12 月.

8.4 向量 ARMA 模型

一元 ARMA 模型也可以推广到处理向量时间序列情形，所得的模型称为 VARMA 模型. 然而，这种推广会遇到一些在 VMA 和 VAR 模型中不会出现的新问题. 其中一个问题是识别问题. 与一元 ARMA 模型不同，VARMA 模型并不是唯一定义的. 例如，VMA(1) 模型：

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

与 VAR(1)模型

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

是相同的. 这两个模型的等价性可以通过检验它们的分量模型很容易地看出. 对 VMA(1)模型，我们有

$$r_{1t} = a_{1t} - 2a_{2,t-1} \quad r_{2t} = a_{2t}$$

322

对这个 VAR(1)模型，方程为

$$r_{1t} + 2r_{2,t-1} = a_{1t} \quad r_{2t} = a_{2t}$$

由 r_{2t}所服从的模型，我们有 $r_{2,t-1}=a_{2,t-1}$. 因此，r_{1t}所服从的模型是相同的. 这种类型的识别问题并没有坏处的，因为在实际应用中，两个模型都可以使用.

另一种类型的识别问题比较麻烦. 考虑 VARMA(1, 1)模型：

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

这个模型与 VARMA(1, 1)模型

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2+\eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} r_{1,t-1} \\ r_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & \eta \\ 0 & \omega \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}$$

对任何非零的 ω 和 η 是相同的. 在这个特例中，等价性之所以成立是因为在两个模型中都有 $r_{2t}=a_{2t}$. 在第二个模型中，参数 ω 和 η 对系统的 AR 部分和 MA 部分的影响互相抵消了. 这一个识别问题是严重的，因为没有恰当的限制，数据的向量 ARMA(1, 1)模型的似然函数并不是唯一定义的，导致了类似于回归分析中的精确多重共线性的问题. 即使没有一个分量是白噪声序列，在一个向量模型中这种类型的识别问题也会发生.

这两个例子引出了推广到 VARMA 模型时涉及到的新问题. 这样，对一个给定的数据集合建立 VARMA 模型要求注意一些问题. 在时间序列文献中，已经提出了结构指定(structural specification)方法来克服这种识别问题，见 Tiao 和 Tsay(1989)，Tsay(1991) 以及里面的参考文献. 我们这里并不讨论结构指定的细节，因为在大多数金融应用中，VAR 和 VMA 模型就足够了. 当使用 VARMA 模型时，只关心低阶的模型(如 VARMA(1, 1)或 VARMA(2, 1)模

型)，尤其是当涉及到的时间序列是非季节性模型时.

VARMA(p，q) 模型可以写成

$$\boldsymbol{\Phi}(B)\boldsymbol{r}_t = \phi_0 + \boldsymbol{\Theta}(B)\boldsymbol{a}_t$$

其中 $\boldsymbol{\Phi}(B)=\boldsymbol{I}-\boldsymbol{\Phi}_1 B-\cdots-\boldsymbol{\Phi}_p B^p$ 和 $\boldsymbol{\Theta}(B)=\boldsymbol{I}-\boldsymbol{\Theta}_1 B-\cdots-\boldsymbol{\Theta}_q B^q$ 是两个 $k\times k$ 矩阵多项式. 我们假定这两个矩阵多项式没有左公共因子，否则模型可以简化. r_t 弱平稳的充要条件与具有矩阵多项式 $\boldsymbol{\Phi}(B)$的 VAR(p)模型是相同的. 对 $v>0$，系数矩阵 $\boldsymbol{\Phi}_v$ 和 $\boldsymbol{\Theta}_v$ 的第(i，j)个元素度量的分别是 r_{1t}对 $r_{j,t-v}$和 $a_{j,t-v}$的线性依赖. 如果对所有的 AR 和 MA 系数矩阵，都有第(i，j)个元素为 0，则 r_{it}并不依赖于 r_{jt}的延迟值. 然而，在 VARMA 模型中反过来并不成立. 换句话说，即使 r_{it}并不依赖于 r_{jt}的任何延迟值，AR 和 MA 矩阵的第(i，j)个位置也可以存在非 0 系数.

323

为了说明上述问题，考虑下面的二元模型：

$$\begin{bmatrix}\Phi_{11}(B) & \Phi_{12}(B)\\ \Phi_{21}(B) & \Phi_{22}(B)\end{bmatrix}\begin{bmatrix}r_{1t}\\ r_{2t}\end{bmatrix}=\begin{bmatrix}\Theta_{11}(B) & \Theta_{12}(B)\\ \Theta_{21}(B) & \Theta_{22}(B)\end{bmatrix}\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}$$

这里从 r_{1t}到 r_{2t}存在单向动态关系的充要条件是

$$\Phi_{22}(B)\Theta_{12}(B)-\Phi_{12}(B)\Theta_{22}(B)=0$$

但是

$$\Phi_{11}(B)\Theta_{21}(B)-\Phi_{21}(B)\Theta_{11}(B)\neq 0 \tag{8.28}$$

这些条件可以如下得到，令

$$\Omega(B)=|\boldsymbol{\Phi}(B)|=\Phi_{11}(B)\Phi_{22}(B)-\Phi_{12}(B)\Phi_{21}(B)$$

为 AR 矩阵多项式的行列式，将模型左乘以矩阵$\begin{bmatrix}\Phi_{22}(B) & -\Phi_{12}(B)\\ -\Phi_{21}(B) & \Phi_{11}(B)\end{bmatrix}$，可以将这个二元模型改写为：

$$\begin{aligned}&\boldsymbol{\Omega}(B)\begin{bmatrix}r_{1t}\\ r_{2t}\end{bmatrix}\\ &=\begin{bmatrix}\Phi_{22}(B)\Theta_{11}(B)-\Phi_{12}(B)\Theta_{21}(B) & \Phi_{22}(B)\Theta_{12}(B)-\Phi_{12}(B)\Theta_{22}(B)\\ \Phi_{11}(B)\Theta_{21}(B)-\Phi_{21}(B)\Theta_{11}(B) & \Phi_{11}(B)\Theta_{22}(B)-\Phi_{21}(B)\Theta_{12}(B)\end{bmatrix}\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}.\end{aligned}$$

考虑 r_{1t}的方程. 方程组(8.28)的第一个条件表示 r_{1t}并不依赖于a_{2t}或r_{2t}的任何过去值. 对 r_{2t}的方程，方程组(8.28)的第二个条件隐含了 r_{2t}确实依赖于 a_{1t}的某些过去值. 根据方程组(8.28)，$\Theta_{12}(B)=\Phi_{12}(B)=0$ 是从 r_{1t}到 r_{2t}存在单向关系的充分但不必要条件.

VARMA 模型的估计可以通过条件或精确最大似然方法来实现. 对拟合模型的残差序列，$Q_k(m)$估计量仍然适用，但是它的渐近 χ^2 分布的自由度为 k^2m-g，这里 g 是 MA 和 AR 系数矩阵中待估参数的个数.

324

例 8.6 为了说明 VARMA 的建模，我们考虑美国的两个月利率序列. 第一个序列是期限为 1 年的国库券利率，第二个序列是期限为 3 年的国库券利率. 数据来自于 St Louis 的联邦储

备银行(FRB)，样本区间是从 1953 年 4 月至 2001 年 1 月，共 574 个观测值. 为了确保利率为正，我们分析对数序列. 图 8-5 显示了两个对数利率序列的时间图. 实线表示 1 年期利率. 这两个序列在样本期间内的运动是非常靠近的.

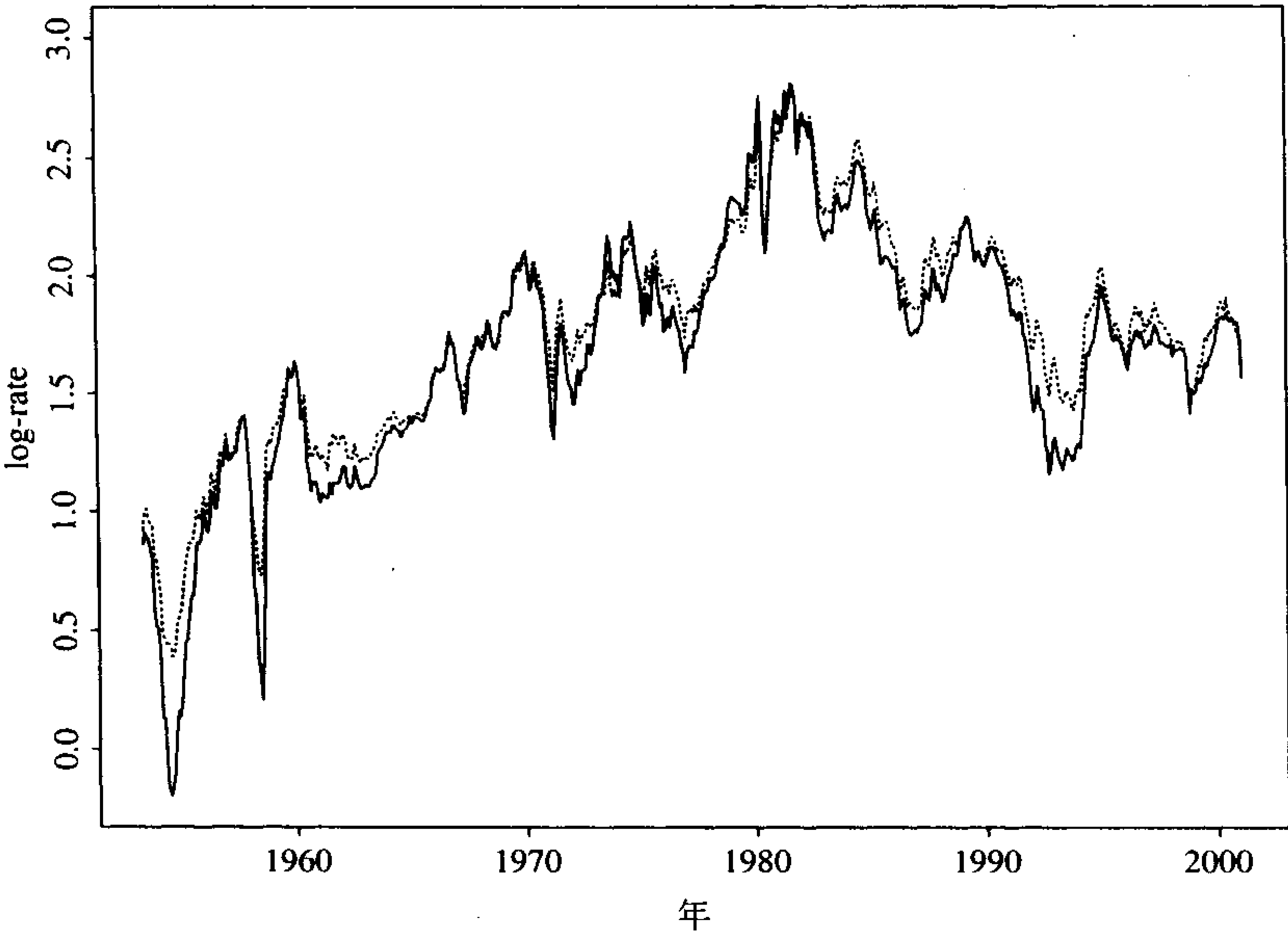
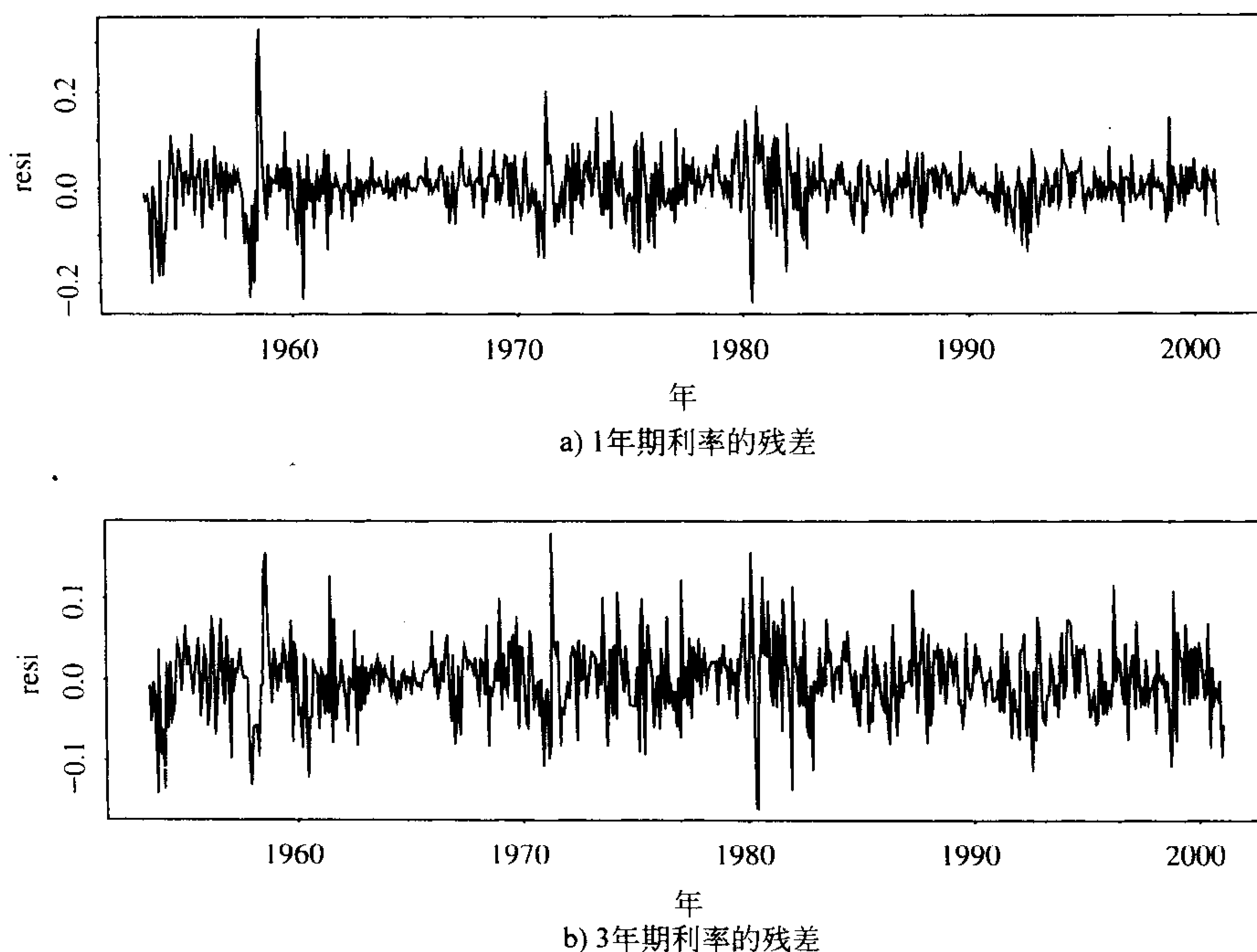


图 8-5　从 1953 年 4 月到 2001 年 1 月美国月利率对数的时间图. 实线为 1 年期的国库券月利率对数，虚线则对应 3 年期

$M(i)$统计量和 AIC 准则对数据指定了一个 VAR(4)模型. 然而，我们采用一个 VARMA(2, 1)模型，因为这两个模型提供了类似的拟合. 表 8-6 给出了由精确似然方法得到的 VARMA(2, 1)模型的参数估计. 我们消除了不显著的参数，重新估计了简化的模型. 拟合模型的残差序列具有轻微的序列相关，且在延迟 7 和 11 处有交叉相关性. 图 8-6 给出了残差图，可见存在一些异常数据点. 模型可以进一步改善，但是，它看上去已经很好地捕获了数据的动态结构.

表 8-6　对美国两个月利率的 VARMA(2, 1)模型的参数估计

参数	$\boldsymbol{\Phi}_1$		$\boldsymbol{\Phi}_2$		$\boldsymbol{\phi}_0$	$\boldsymbol{\Theta}_1$		$\Sigma\times10^3$	
估计	1.57	−0.54	−0.60	0.56	0.020	0.60	−1.17	3.58	2.50
		0.99			0.025		−0.47	2.50	2.19
标准	0.10	0.16	0.09	0.15	0.013	0.11	0.18		
		0.01			0.011		0.04		



a) 1年期利率的残差

b) 3年期利率的残差

图 8-6 对美国的两个月利率序列拟合 VARMA(2, 1)模型的残差图

最后的 VARMA(2.1)模型给出了数据的一些有趣的特征. 第一，利率序列是高度同步相关的. 同步相关系数为 $2.5/\sqrt{3.58\times 2.19}=0.893$. 第二，从 3 年期利率到 1 年期利率有单向线性关系，因为所有 AR 和 MA 矩阵的第(2, 1)个元素都为 0，但某些第(2, 1)个元素不是 0. 事实上，表 8-6 中的模型说明：

$$r_{3t}=0.025+0.99r_{3,t-1}+a_{3t}+0.47a_{3,t-1}$$

$$r_{1t}=0.020+1.57r_{1,t-1}-0.60r_{1,t-2}-0.54r_{3,t-1}+0.56r_{3,t-2}$$
$$+a_{1t}-0.60a_{1,t-1}+1.17a_{3,t-1}$$

325 ~ 326

其中 r_{it} 为 i 年期利率的对数序列，a_{it} 为相应的抖动序列. 因此，3 年期利率并不依赖于 1 年期利率的过去值，但是 1 年期利率依赖于 3 年期利率的过去值. 第三，这两个利率序列看上去似乎是单位根非平稳的，利用向后推移算子 B，模型可以近似改写为：

$$(1-B)r_{3t}=0.025+(1+0.47B)a_{3t}$$

$$(1-B)(1-0.6B)r_{1t}=0.02-0.55B(1-B)r_{3,t}+(1-0.6B)a_{1t}+1.17Ba_{3,t}$$

分量的边际模型

给定 $\boldsymbol{r}_t$ 的一个向量模型，组成部分 r_{it} 的隐含的一元模型就是边际模型. 对一个 k 维 ARMA(p, q)模型，边际模型是 ARMA[kp, $(k-1)p+q$]. 这个结果可以通过两步得到. 首先，VMA(q)的边际模型是一元 MA(q). 假定 $\boldsymbol{r}_t$ 是一个 MA(q)过程. 因为 $\boldsymbol{r}_t$ 的交叉相关矩阵

在延迟 q 步截尾(即对 $l>q$，有 $\boldsymbol{\rho}_l=\boldsymbol{0}$)，$r_{it}$的 ACF 在超出延迟 q 后是 0. 因此，r_{it}是一个 MA 过程，并且它的一元模型具有形式 $r_{it}=\theta_{i,0}+\sum_{j=1}^{q}\theta_{i,j}b_{i,t-j}$，这里$\{b_{it}\}$是不相关的随机变量序列，均值为 0，方差为 σ_{ib}^2. 参数 $\theta_{i,j}$和σ_{ib}是$\boldsymbol{r}_t$ 的 VMA 模型参数的函数.

第二步要得到的结果是对角化 VARMA(p, q)模型的 AR 矩阵多项式. 为了更好地说明，考虑二元 AR(1)模型：

$$\begin{bmatrix}1-\Phi_{11}(B) & -\Phi_{12}(B)\\ -\Phi_{21}(B) & 1-\Phi_{22}(B)\end{bmatrix}\begin{bmatrix}r_{1t}\\ r_{2t}\end{bmatrix}=\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}.$$

左乘矩阵多项式$\begin{bmatrix}1-\Phi_{22}(B) & \Phi_{12}(B)\\ \Phi_{21}(B) & 1-\Phi_{11}(B)\end{bmatrix}$，我们得到：

$$[(1-\Phi_{11}B)(1-\Phi_{22}B)-\Phi_{12}\Phi_{22}B^2]\begin{bmatrix}r_{1t}\\ r_{2t}\end{bmatrix}=\begin{bmatrix}1-\Phi_{22}(B) & \Phi_{12}(B)\\ \Phi_{21}(B) & 1-\Phi_{11}(B)\end{bmatrix}\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}$$

前面方程的左边表明 r_{it}的一元 AR 多项式是 2 阶的，而方程的右面具有 VMA(1)的形式. 利用第一步中 VMA 模型的结果，我们证明了 r_{it}的一元模型是 ARMA(2, 1)的. 这个方法很容易推广到 k 维 VAR(1)模型，其边际模型为 ARMA(k, $k-1$). 更一般地，对一个 k 维 VAR(p)模型，其边际模型为 ARMA[kp, $(k-1)p$]. VARMA 模型的结果可直接从 VMA 与 VAR 模型的结果中得到.

阶[kp, $(k-1)p+q$]是边际模型的最大阶(即上界). r_{it}的真实边际阶可能会更低. 327

8.5 单位根非平稳性与协整

当对几个单位根非平稳时间序列联合建模时，可能遇到协整的情况. 考虑二元 ARMA(1, 1)模型.

$$\begin{bmatrix}x_{1t}\\ x_{2t}\end{bmatrix}-\begin{bmatrix}0.5 & -1.0\\ -0.25 & 0.5\end{bmatrix}\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix}=\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}-\begin{bmatrix}0.2 & -0.4\\ -0.1 & 0.2\end{bmatrix}\begin{bmatrix}a_{1,t-1}\\ a_{2,t-1}\end{bmatrix} \tag{8.29}$$

其中抖动 a_t 的协方差矩阵Σ是正定的. 这并不是一个弱平稳模型，因为 AR 系数矩阵的两个特征值分别为 0 和 1. 图 8-7 给出了$\Sigma=\boldsymbol{I}$ 时模型的 200 个模拟数据点的时间图；而图 8-8 给出了两个分量序列 x_{it}的样本自相关系数，很容易看出，这两个序列具有高度自相关而且表现单位根非平稳性. x_t 的两个边际模型的确是单位根非平稳的. 将模型改写为

$$\begin{bmatrix}1-0.5B & B\\ 0.25B & 1-0.5B\end{bmatrix}\begin{bmatrix}x_{1t}\\ x_{2t}\end{bmatrix}=\begin{bmatrix}1-0.2B & 0.4B\\ 0.1B & 1-0.2B\end{bmatrix}\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}$$

将上述方程左乘以$\begin{bmatrix}1-0.5B & -B\\ -0.25B & 1-0.5B\end{bmatrix}$，得到

$$\begin{bmatrix}1-B & 0\\ 0 & 1-B\end{bmatrix}\begin{bmatrix}x_{1t}\\ x_{2t}\end{bmatrix}=\begin{bmatrix}1-0.7B & -0.6B\\ -0.15B & 1-0.7B\end{bmatrix}\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}.$$

因此，模型的每个分量 x_{it} 都不是单位根平稳的，且服从一个 ARIMA(0，1，1)模型.

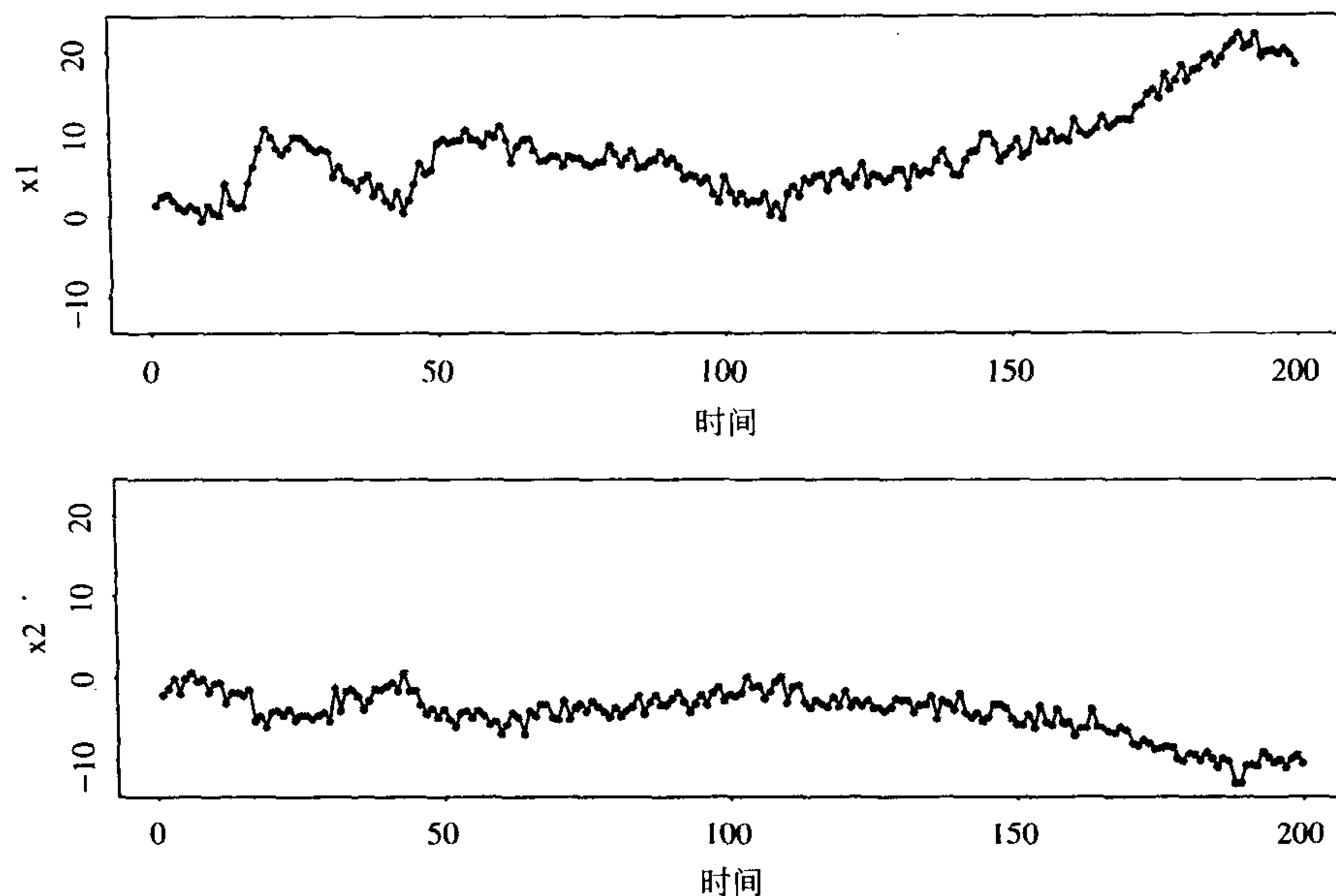


图 8-7　当抖动的协方差矩阵为单位阵时的式(8.29)所定义模型的模拟序列的时间图

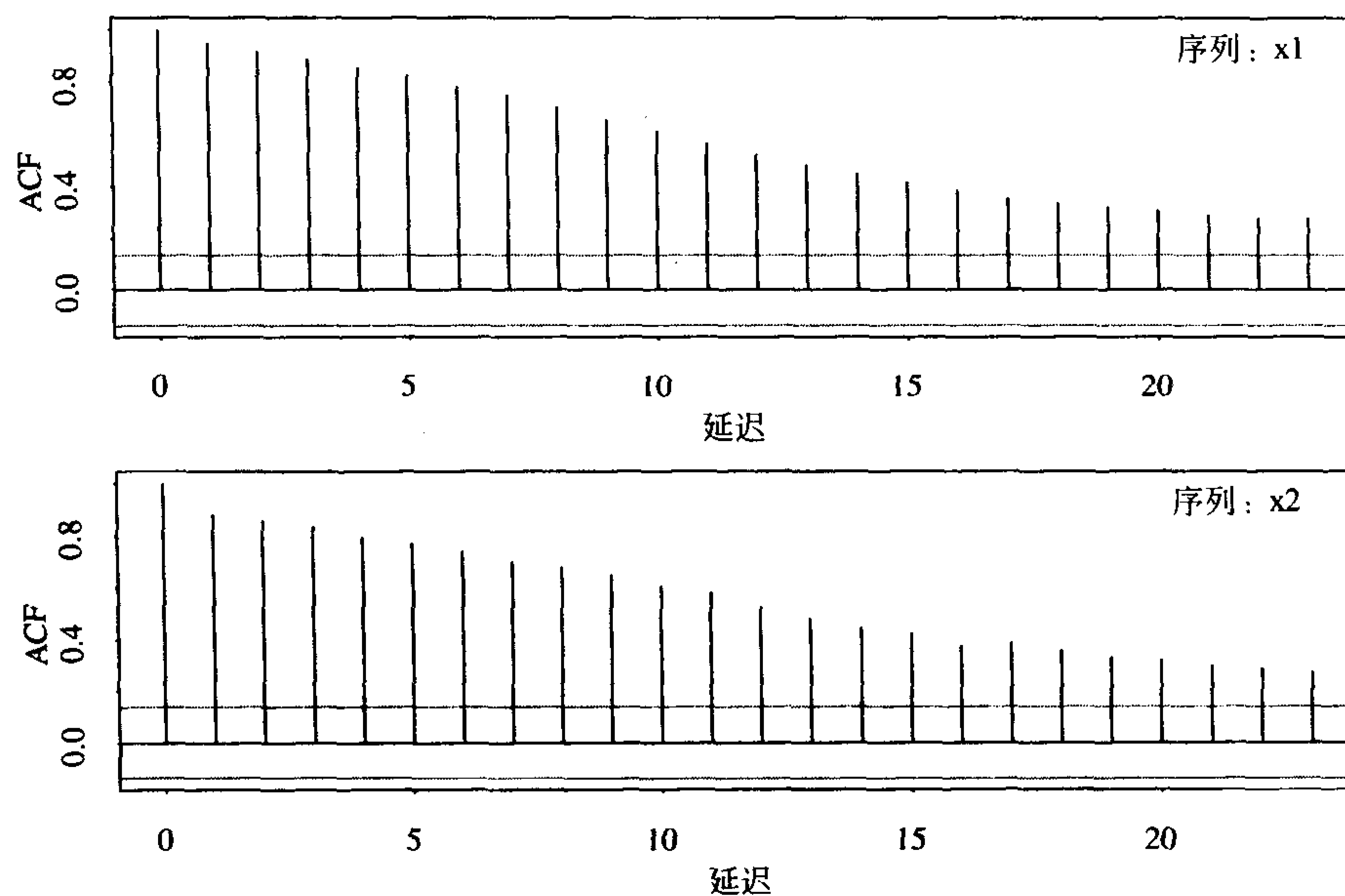


图 8-8　当抖动的协方差矩阵为单位阵时的式(8.29)所定义模型的两个模拟分量序列的样本自相关函数图. 观测值为 200 个

然而，我们可以考虑如下一个线性变换，定义

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{x}_t$$

$$\begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 & -2.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \equiv \boldsymbol{L}\boldsymbol{a}_t$$

328～329

变换序列 $\boldsymbol{y}_t$ 的 VARMA 模型可以如下得到：

$$\begin{aligned} \boldsymbol{L}\boldsymbol{x}_t &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{x}_{t-1} + \boldsymbol{L}\boldsymbol{a}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{a}_{t-1} \\ &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}^{-1}(\boldsymbol{L}\boldsymbol{x}_{t-1}) + \boldsymbol{b}_t - \boldsymbol{L}\boldsymbol{\Theta}\boldsymbol{L}^{-1}\boldsymbol{b}_{t-1} \end{aligned}$$

这样，$\boldsymbol{y}_t$ 的模型为

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} - \begin{bmatrix} 1.0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} = \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} - \begin{bmatrix} 0.4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_{1,t-1} \\ b_{2,t-1} \end{bmatrix} \tag{8.30}$$

由此模型看出

(a) y_{1t} 与 y_{2t} 是分离序列，其同步相关系数等于抖动 b_{1t} 与 b_{2t} 之间的同步相关系数；

(b) y_{1t} 服从一个一元 ARIMA(0，1，1)模型；

(c) y_{2t} 是一个白噪声序列(即 $y_{2t}=b_{2t}$).

特别地，方程(8.30)的模型表明系统中只有一个单位根. 因此 x_{1t} 与 x_{2t} 的单位根是由 y_{1t} 的单位根引入的.

x_{1t} 与 x_{2t} 都是单位根非平稳的，但在向量序列中只存在一个单位根，这种现象在计量经济和时间序列文献中称为协整(co-integration). 另一个定义协整的方式着眼于单位根非平稳序列的线性变换. 对模型(8.29)的模拟例子，变换表明线性组合 $y_{2t}=0.5x_{1t}+x_{2t}$ 没有单位根. 如果(a) x_{1t} 和 x_{2t} 都不是单位根平稳的；(b)它们的一个线性组合具有单位根平稳性，则定义 x_{1t} 与 x_{2t} 是协整的.

一般而言，对一个 k 维单位根非平稳时间序列，如果系统中低于 k 个单位根，则协整存在. 令 h 表示 k 维序列 $\boldsymbol{x}_t$ 中单位根的数量. 如果 $0<h<k$，则存在协整. 而且，$k-h$ 为协整因子的数目. 换言之，协整因子的个数为具有单位根平稳性的不同线性组合的个数，这种线性组合称为协整向量. 对前面模拟的例子，$y_{2t}=(0.5, 1)\boldsymbol{x}_t$，因而 $(0.5, 1)'$ 是系统的一个协整向量. 更多关于协整和协整检验的讨论，参见 Box 和 Tiao(1977)，Engle 和 Granger (1987)，Stock 和 Watson (1988) 以及 Johansen(1989).

协整的概念是很有趣的，在文献中已经引起了大量关注. 然而，实际应用中对协整的检验有些困难. 困难的主要原因是协整检验忽视了分量序列的尺度效应. 感兴趣的读者可以参考 Cochrane(1988)，Tiao，Tsay 和 Wang(1993)，这些文献中有进一步的讨论.

尽管我对协整检验的实际价值感到怀疑，然而协整的思想与金融研究是高度相关的. 例如，考虑 Finnish Nokia 公司的股票. 它在 Helsinki 股市上的价格必须与纽约证券交易所中它的美国信托收据(American Depositary Receipts)的价格联动；否则对投资者就存在套利机会.

330

如果股价有单位根，则两个价格序列一定是协整的．实际中，在调整交易成本和汇率风险之后，就存在这样的协整．稍后在 8.6 节中我们讨论这样的问题．

误差修正形式

因为在协整系统中，单位根非平稳分量的个数多于单位根的个数，所以对于单个分量差分所得到的平稳性结果是差分过度的．过度差分导致了 MA 矩阵多项式中的单位根问题，反过来可能在参数估计中遇到困难．如果 MA 矩阵多项式包含单位根，则说向量时间序列是不可逆的．

Engle 和 Granger(1987)讨论了一个协整系统的误差修正表示，它克服了不可逆的 VARMA 模型估计中的困难．考虑方程(8.29)的协整系统．令$\nabla \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$是差分后的序列．将方程的两边减去 $\boldsymbol{x}_{t-1}$，我们得到对$\nabla \boldsymbol{x}_t$ 的一个模型为

$$\begin{bmatrix}\nabla x_{1t}\\ \nabla x_{2t}\end{bmatrix}=\begin{bmatrix}-0.5 & -1.0\\ -0.25 & -0.5\end{bmatrix}\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix}+\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}-\begin{bmatrix}0.2 & -0.4\\ -0.1 & 0.2\end{bmatrix}\begin{bmatrix}a_{1,t-1}\\ a_{2,t-1}\end{bmatrix}$$

$$=\begin{bmatrix}-1\\ -0.5\end{bmatrix}[0.5 \quad 1.0]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix}+\begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}-\begin{bmatrix}0.2 & -0.4\\ -0.1 & 0.2\end{bmatrix}\begin{bmatrix}a_{1,t-1}\\ a_{2,t-1}\end{bmatrix}$$

这是一个平稳模型，因为$\nabla \boldsymbol{x}_t$ 与 $[0.5,\quad 1.0]\boldsymbol{x}_t = y_{2t}$都是单位根平稳的．因为上述方程的右端用到了 $\boldsymbol{x}_{t-1}$，所以 MA 矩阵多项式同未差分前一样，从而模型不会遇到不可逆问题．这个公式称为$\nabla \boldsymbol{x}_t$ 的误差修正模型．它可以扩展到一般的协整 VARMA 模型．对一个具有 m 个协整因子的协整 VARMA(p, q)模型，其误差修正表示为

$$\nabla \boldsymbol{x}_t = \boldsymbol{\alpha\beta x}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \ \nabla \boldsymbol{x}_{t-i} + \boldsymbol{a}_t - \sum_{j=1}^{q} \boldsymbol{\Theta}_j \boldsymbol{a}_{t-j} \tag{8.31}$$

这里 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 分别是 $k\times m$ 和 $m\times k$ 的满秩矩阵．AR 系数矩阵 $\boldsymbol{\Phi}_i^*$ 是原始系数矩阵 $\boldsymbol{\Phi}_j$ 的函数；见后面的注释．时间序列 $\boldsymbol{\beta x}_t$ 是单位根平稳的，并且 $\boldsymbol{\beta}$ 的行向量为 $\boldsymbol{x}_t$ 的协整向量．

在误差修正表示中出现平稳序列 $\boldsymbol{\beta x}_{t-1}$是很自然的，可以认为是过度差分系统$\nabla \boldsymbol{x}_t$ 中的一个“弥补”项．$\boldsymbol{\beta x}_{t-1}$平稳性的证明如下：单位根时间序列理论证明了单位根非平稳序列与平稳序列之间的样本相关系数在样本量趋于无穷时收敛到 0；见 Tsay 和 Tiao(1990)及里面的参考文献．在一个误差修正表示中，$\boldsymbol{x}_{t-1}$是单位根非平稳的，但是$\nabla \boldsymbol{x}_t$ 是平稳的，因此，将$\nabla \boldsymbol{x}_t$ 与 $\boldsymbol{x}_{t-1}$有意义地联系起来的唯一方式是通过平稳序列 $\boldsymbol{\beta x}_{t-1}$．

331

注释：方程(8.31)中误差修正模型的 AR 系数矩阵由下式给出

$$\begin{aligned}
\boldsymbol{\Phi}_{p-1}^* &= -\boldsymbol{\Phi}_p\\
\boldsymbol{\Phi}_{p-2}^* &= -\boldsymbol{\Phi}_{p-1} - \boldsymbol{\Phi}_p\\
\vdots &= \vdots\\
\boldsymbol{\Phi}_1^* &= -\boldsymbol{\Phi}_2 - \cdots - \boldsymbol{\Phi}_p\\
\boldsymbol{\alpha\beta} &= \boldsymbol{\Phi}_p + \boldsymbol{\Phi}_{p-1} + \cdots + \boldsymbol{\Phi}_1 - \boldsymbol{I} = -\boldsymbol{\Phi}(1)
\end{aligned}$$

注释：我们对协整的讨论假定所有的单位根重数为 1，但是这个概念可以扩展为单位根为多重的情形．如果协整因子的个数 m 给定，则方程(8.31)的误差修正模型仍然

可以通过似然方法来估计. 最后，有许多方式可以构造误差修正表示. 事实上，在方程(8.31)中，只要对 AR 系数矩阵 $\boldsymbol{\Phi}_i^*$ 进行某些修正，可以用任何 $\boldsymbol{\alpha\beta x}_{t-v}(1\leqslant v\leqslant p)$.

8.6 门限协整与套利

本节中，我们主要讨论在指数交易中利用多元时间序列方法探测套利机会. 我们还指出，与协整思想相结合，第 4 章中简单的一元非线性模型可以自然地推广到多元情形.

我们的研究考虑标准普尔 500 指数期货的价格与现金市场上以该指数为标的的股份价格之间的关系. 令 $f_{t,l}$表示到期时间为 l 的指数期货在时刻 t 的对数价格，s_t 表示现金市场上的以指数为标的的股份在 t 时刻的对数价格. 金融文献中，实现成本模型(cost-of-carry model)的一个版本规定：

$$f_{t,l}-s_t=(r_{t,l}-q_{t,l})(l-t)+z_t^* \tag{8.32}$$

这里 $r_{t,l}$是无风险利率，$q_{t,l}$是关于 t 时刻现金价格的红利收益，$(l-t)$是期货合约的到期时间；见 Brenner 和 Kroner(1995)，Dwyer，Locke 和 Yu(1996) 及其参考文献.

模型(8.32)中的过程 z_t^* 必须是单位根平稳的；否则就存在持续的套利机会. 这里的套利交易包括，当期货合约在到期日以前的对数价格偏离到大于随时间持有指数的成本时，同时购买(卖空)证券指数；并卖出(购买)指数期货. 在 z_t^* 的弱平稳假定下，为了有利可图，z_t^* 的模一定超过一个由交易成本和其他的经济因素和风险因素决定的一个特定值. 332

通常认为标准普尔 500 指数的 $f_{t,l}$与 s_t 序列都包含一个单位根，但是方程(8.32)表明它们在调整利率和红利收益的影响之后是协整的. 调整后的协整向量为(1，−1)，协整序列是 z_t^*. 因此，应该利用误差修正形式来对收益率序列$\boldsymbol{r}_t=(\nabla f_t,\ \nabla s_t)'$建模，这里$\nabla f_t=f_{t,l}-f_{t-1,l}$，$\nabla s_t=s_t-s_{t-1}$，并且为了记号的简便，我们将到期时间 l 从∇f_t 的下标中去掉了.

8.6.1 多元门限模型

实际中，套利交易影响市场动态，从而，依赖于套利交易的存在与否，$\boldsymbol{r}_t$ 的模型可能随时间变化. 因此，前面的讨论自然引出了下述模型

$$\boldsymbol{r}_t=\begin{cases}\boldsymbol{c}_1+\sum_{i=1}^{p}\boldsymbol{\Phi}_i^{(1)}\boldsymbol{r}_{t-i}+\boldsymbol{\beta}_1 z_{t-1}+\boldsymbol{a}_t^{(1)}, & 若\ z_{t-1}\leqslant\gamma_1,\\ \boldsymbol{c}_2+\sum_{i=1}^{p}\boldsymbol{\Phi}_i^{(2)}\boldsymbol{r}_{t-i}+\boldsymbol{\beta}_2 z_{t-1}+\boldsymbol{a}_t^{(2)}, & 若\ \gamma_1<z_{t-1}\leqslant\gamma_2,\\ \boldsymbol{c}_3+\sum_{i=1}^{p}\boldsymbol{\Phi}_i^{(3)}\boldsymbol{r}_{t-i}+\boldsymbol{\beta}_3 z_{t-1}+\boldsymbol{a}_t^{(3)}, & 若\ \gamma_2<z_{t-1},\end{cases} \tag{8.33}$$

其中 $z_t=100z_t^*$，$\gamma_1<0<\gamma_2$ 是两个实数，$\{\boldsymbol{a}_t^{(i)}\}$是彼此独立的 2 维白噪声序列. 这里我们用 $z_t=100z_t^*$，因为 z_t^* 的实际值相对较小.

方程(8.33) 中的模型称为三体制的多元门限模型. 两个实数 γ_1 和 γ_2 是门限，z_{t-1}是门限变量. 门限变量 z_{t-1}是由数据支持的；见 Tsay(1998). 一般考虑 $d\in\{1,\ \cdots,\ d_0\}$，可以选择

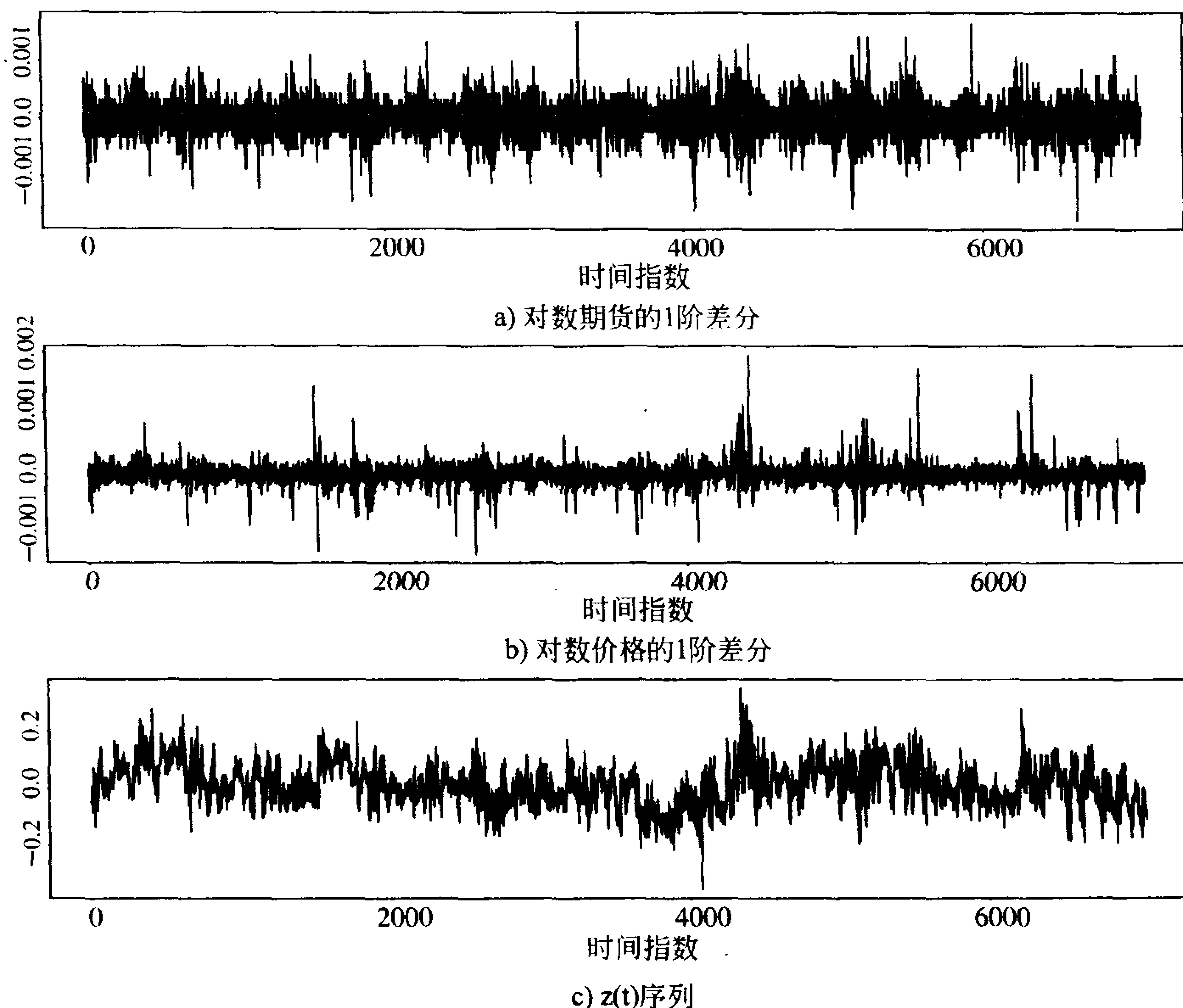
z_{t-d}作为一个门限变量，其中 d_0 是一个预先指定的正整数.

模型(8.33)是第 4 章中门限自回归模型的推广，它也是方程(8.31)中误差修正模型的推广. 如前所提到的，只有当 z_t^*(或等价地，z_t)的模相当大时，套利交易才是有利可图的. 因此，模型(8.33)只有在体制 1 和体制 3 会有套利交易发生. 这样，体制 2 中的 $f_{t,l}$与 s_t 间的动态关系主要是由正常市场力量决定的，从而两个序列或多或少类似于一个随机游走. 换句话说，中间体制的两个对数价格应该不受套利的影响，从而也不受协整的限制. 从经济计量的观点看，这意味着中间体制 β_2 的估计应该是不显著的.

333 总之，我们期望期货的对数价格与现金市场上的证券指数的对数之间的协整效应在体制 1 与体制 3 中都是显著的，但是在体制 2 中不显著. 这种现象称为门限协整，见 Balke 和 Fomby(1997).

8.6.2 数据

在下面的实例研究中使用的数据是标准普尔 500 指数 1993 年 5 月的日内交易数据与它在芝加哥商品交易所 6 月份期货合约，见 Forbes，Kalb 和 Kofman(1999)，他们利用这个数据构造了一个具有 7060 个观测值的每分钟的二元价格序列. 为了避免异常收益率的过度影响，我将十个极值(每边 5 个)用与它们最近的两个值的简单平均来代替. 这一步并不影响分析的定性结论，但是可能影响数据的条件异方差性. 为了简便，我们在研究中不考虑条件异方差性. 图 8-9是指数期货和现金价格的对数收益率的时间图，与模型(8.32)相应的门限变量$z_t=100z_t^*$.



334 图 8-9 1993 年 5 月标准普尔 500 指数期货和现金价格的 1 分钟对数收益率的时间图及相关的门限变量

8.6.3 估计

方程(8.33)中多元门限模型的一个正式识别包括选择门限变量，决定体制个数，以及对每个体制选择阶 p. 感兴趣的读者可以参考 Tsay(1998)，Forbes，Kalb 和 Kofman(1999). 可以利用一些信息准则(如 Akaike 信息准则(AIC)或残差的平方和)来估计门限 γ_1 和 γ_2. 假定 $p=8$，$d\in\{1, 2, 3, 4\}$，$\gamma_1\in[-0.15, -0.02]$，$\gamma_2\in[0.025, 0.145]$. 并利用格点搜索方法(每个区间内有 300 个点，有 2 个区间)，AIC 选择 z_{t-1}作为门限变量时，门限为$\hat{\gamma}_1=-0.0226$，$\hat{\gamma}_2=0.0377$. 参数估计的细节见表 8-7.

表 8-7

	体制 1		体制 2		体制 3	
	∇f_t	∇s_t	∇f_t	∇s_t	∇f_t	∇s_t
ϕ_0	0.000 02	0.000 05	0.000 00	0.000 00	−0.000 01	−0.000 05
t	(1.47)	(7.64)	(−0.07)	(0.53)	(−0.74)	(−6.37)
∇f_{t-1}	−0.084 68	0.070 98	−0.038 61	0.040 37	−0.041 02	0.023 05
t	(−3.83)	(6.15)	(−1.53)	(3.98)	(−1.72)	(1.96)
∇f_{t-2}	−0.004 50	0.158 99	0.044 78	0.086 21	−0.020 69	0.098 98
t	(−0.20)	(13.36)	(1.85)	(8.88)	(−0.87)	(8.45)
∇f_{t-3}	0.022 74	0.119 11	0.072 51	0.097 52	0.003 65	0.084 55
t	(0.95)	(9.53)	(3.08)	(10.32)	(0.15)	(7.02)
∇f_{t-4}	0.024 29	0.081 41	0.014 18	0.068 27	−0.027 59	0.076 99
t	(0.99)	(6.35)	(0.60)	(7.24)	(−1.13)	(6.37)
∇f_{t-5}	0.003 40	0.089 36	0.011 85	0.048 31	−0.006 38	0.050 04
t	(0.14)	(7.10)	(0.51)	(5.13)	(−0.26)	(4.07)
∇f_{t-6}	0.000 98	0.072 91	0.012 51	0.035 80	−0.039 41	0.026 15
t	(0.04)	(5.64)	(0.54)	(3.84)	(−1.62)	(2.18)
∇f_{t-7}	−0.003 72	0.052 01	0.029 89	0.048 37	−0.020 31	0.022 93
t	(−0.15)	(4.01)	(1.34)	(5.42)	(−0.85)	(1.95)
∇f_{t-8}	0.000 43	0.009 54	0.018 12	0.021 96	−0.044 22	0.004 62
t	(0.02)	(0.76)	(0.85)	(2.57)	(−1.90)	(0.40)
∇s_{t-1}	−0.084 19	0.002 64	−0.076 18	−0.056 33	0.066 64	0.111 43
t	(−2.01)	(0.12)	(−1.70)	(−3.14)	(1.49)	(5.05)
∇s_{t-2}	−0.051 03	0.002 56	−0.109 20	−0.015 21	0.040 99	−0.011 79
t	(−1.18)	(0.11)	(−2.59)	(−0.90)	(0.92)	(−0.53)
∇s_{t-3}	0.072 75	−0.036 31	−0.005 04	0.011 74	−0.019 48	−0.018 29
t	(1.65)	(−1.58)	(−0.12)	(0.71)	(−0.44)	(−0.84)
∇s_{t-4}	0.047 06	0.014 38	0.027 51	0.014 90	0.016 46	0.003 67
t	(1.03)	(0.60)	(0.71)	(0.96)	(0.37)	(0.17)

（续）

	体制 1		体制 2		体制 3	
	∇f_t	∇s_t	∇f_t	∇s_t	∇f_t	∇s_t
∇s_{t-5}	0.081 18	0.021 11	0.039 43	0.023 30	−0.034 30	−0.004 62
t	(1.77)	(0.88)	(0.97)	(1.43)	(−0.83)	(−0.23)
∇s_{t-6}	0.043 90	0.045 69	0.016 90	0.019 19	0.060 84	−0.003 92
t	(0.96)	(1.92)	(0.44)	(1.25)	(1.45)	(−0.19)
∇s_{t-7}	−0.030 33	0.020 51	−0.086 47	0.002 70	−0.004 91	0.035 97
t	(−0.70)	(0.91)	(−2.09)	(0.16)	(−0.13)	(1.90)
∇s_{t-8}	−0.029 20	0.030 18	0.018 87	−0.002 13	0.000 30	0.021 71
t	(−0.68)	(1.34)	(0.49)	(−0.14)	(−0.01)	(1.14)
z_{t-1}	0.000 24	0.000 97	−0.000 10	0.000 12	0.000 25	0.000 86
t	(1.34)	(10.47)	(−0.30)	(0.86)	(1.41)	(9.75)

注：1993 年 5 月标准普尔 500 指数数据所建方程(8.33)中的多元门限模型参数的最小二乘估计及它们的 t 比．三个体制的数据点数分别为 2 234，2 410 和 2 408.

由表 8-7 我们观测到如下几点：首先，中间体制中$\hat{\beta}_2$ 的 t 比表明，如我们所料，该估计在 5%水平下不显著，证实了当不出现套利机会时，两个对数价格之间没有协整．第二，对所有三个体制，∇f_t 都与∇f_{t-1}负相关，这与第 5 章所讨论的买卖弹性是一致的．第三，指数期货的过去对数收益率看上去比现金价格的过去对数收益率包含更多的信息，因为∇f_{t-i}比∇s_{t-i}具有更显著的 t 比．这是合理的，因为一般来说，期货的流动性更强．关于指数套利的更多信息，参见 Dwyer，Locke 与 Yu(1996)．

8.7 主成分分析

前几节中，我们主要讨论了向量时间序列动态结构的建模问题．在多元时间序列分析中，序列的协方差(或相关系数)结构同等重要．例如，向量收益率序列的协方差结构在组合选择中起着很重要的作用．下面，我们讨论研究向量时间序列的协方差结构时的一些有用的统计方法．

给定一个 k 维随机向量 $\boldsymbol{r}=(r_1, \cdots, r_k)'$，其协方差矩阵为$\Sigma_r$，则主成分分析(principal component analysis，PCA) 关心的是利用 r_i 很少的线性组合来解释Σ_r 的结构．如果 $\boldsymbol{r}$ 表示 k 个资产的月对数收益率，则可用 PCA 来研究这 k 个资产收益率变化的原因．这里关键词是少数，以便使得多元分析可以获得简化．

8.7.1 PCA 理论

335～336

PCA 对 $\boldsymbol{r}$ 的协方差矩阵Σ_r 或相关矩阵 $\boldsymbol{\rho}_r$ 都适用．因为相关矩阵是标准化随机向量 $\boldsymbol{r}^*=\boldsymbol{D}^{-1}\boldsymbol{r}$ 的协方差矩阵，此处 $\boldsymbol{D}$ 是 $\boldsymbol{r}$ 的分量的标准差组成的对角矩阵，所以在我们的理论分析中

使用协方差矩阵. 令 $\boldsymbol{c}_i=(c_{i1}, \cdots, c_{ik})'$表示 k 维向量，这里 $i=1, \cdots, k$. 那么

$$y_i = \boldsymbol{c}'_i\boldsymbol{r}$$

是随机向量 $\boldsymbol{r}$ 的线性组合. 若 $\boldsymbol{r}$ 由 k 支股票的简单收益率组成，则 y_i 是将第 j 支股票加权重 c_{ij} 之后的一个组合的收益率. 因为将 $\boldsymbol{c}_i$ 乘上一个常数并不影响分配到第 j 支股票的权重比例，所以我们将向量 $\boldsymbol{c}_i$ 标准化，使得 $\boldsymbol{c}'_i\boldsymbol{c}_i = \sum_{j=1}^{k} c_{ij}^2 = 1$.

利用随机变量线性组合的性质，我们有

$$\mathrm{Var}(y_i)=\boldsymbol{c}'_i\Sigma_r\boldsymbol{c}_i \qquad i=1, \cdots, k \tag{8.34}$$

$$\mathrm{Cov}(y_i, y_j)=\boldsymbol{c}'_i\Sigma_r\boldsymbol{c}_j \quad i, j=1, \cdots, k \tag{8.35}$$

PCA 的思想就是找到线性组合 $\boldsymbol{c}_i$，对 $i\neq j$，y_i 与 y_j 是不相关的，并且 y_i 的方差尽可能大. 更具体地，

1) $\boldsymbol{r}$ 的第一个主成分是 在 $\boldsymbol{c}'_1\boldsymbol{c}_1=1$ 的限制下，使得 $\mathrm{Var}(y_1)$最大的线性组合 $y_1=\boldsymbol{c}'_1\boldsymbol{r}$;

2) $\boldsymbol{r}$ 的第二个主成分是在 $\boldsymbol{c}'_2\boldsymbol{c}_2=1$ 与 $\mathrm{Cov}(y_2, y_1)=0$ 的限制下，使得 $\mathrm{Var}(y_2)$最大的线性组合 $y_2=\boldsymbol{c}'_2\boldsymbol{r}$;

3) $\boldsymbol{r}$ 的第 i 个主成分是在 $\boldsymbol{c}'_i\boldsymbol{c}_i=1$ 与 $\mathrm{Cov}(y_i, y_j)=0 (j=1, \cdots, i-1)$的限制下，最大化 $\mathrm{Var}(y_i)$的线性组合 $y_i=\boldsymbol{c}'_i\boldsymbol{r}$.

因为Σ_r 的协方差矩阵是非负定的，所以具有谱分解，见附录 A. 令$(\lambda_1, \boldsymbol{e}_1), \cdots, (\lambda_k, \boldsymbol{e}_k)$ 为Σ_r 的特征值-特征向量对，其中 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k\geqslant 0$，则我们有下面的统计结果.

结果：$\boldsymbol{r}$ 的第 i 个主成分是 $y_i = \boldsymbol{e}'_i\boldsymbol{r} = \sum_{j=1}^{k} e_{ij}r_j (i = 1,\cdots,k)$. 而且

$$\mathrm{Var}(y_i)=\boldsymbol{e}'_i\Sigma_r\boldsymbol{e}_i=\lambda_i \qquad i=1, \cdots, k,$$

$$\mathrm{Cov}(y_i, y_j)=\boldsymbol{e}'_i\Sigma_r\boldsymbol{e}_j=0 \quad i\neq j.$$

如果某些特征值 λ_i 是相等的，则对应特征向量 $\boldsymbol{e}_i$ 的选择不是唯一的，从而 y_i 也不是唯一的. 另外，我们有

$$\sum_{i=1}^{k}\mathrm{Var}(r_i) = tr(\Sigma_r) = \sum_{i=1}^{k}\lambda_i = \sum_{i=1}^{k}\mathrm{Var}(y_i) \tag{8.36}$$

337

等式(8.36)说明

$$\frac{\mathrm{Var}(y_i)}{\sum_{i=1}^{k}\mathrm{Var}(r_i)} = \frac{\lambda_i}{\lambda_1+\cdots+\lambda_k}.$$

因此，$\boldsymbol{r}$ 的总方差中由第 i 个主成分解释的比例是Σ_r 的第 i 个特征值占Σ_r 的所有特征值总和的比率. 也可以计算由前 i 个主成分所能解释的总方差的累积比例 $\left(即\left(\sum_{j=1}^{i}\lambda_j\right)\Big/\left(\sum_{j=1}^{k}\lambda_j\right)\right)$. 实际中，可以选择一个较小的 i，使得前面的累积比例很大.

因为 $tr(\rho_r)=k$，所以当采用相关阵来进行主成分分析时，由第 i 个主成分解释的方差比例变为 λ_i/k.

PCA 的一个副产品是Σ_r或$\boldsymbol{\rho}_r$的 0 特征值，表明$\boldsymbol{r}$的分量之间存在精确的线性关系. 例如，如果最小特征值$\lambda_k=0$，则由前面的结果$\mathrm{Var}(y_k)=0$. 因此，$y_k=\sum_{j=1}^{k}e_{kj}r_j$是个常数，从而在$\boldsymbol{r}$中只有$k-1$个随机量. 在这种情形下，$\boldsymbol{r}$的维数可以降低. 由于这个原因，文献中常用 PCA 作为降低维数的工具.

8.7.2 经验的 PCA

应用中，收益率向量$\boldsymbol{r}$的协方差矩阵Σ_r和相关矩阵$\boldsymbol{\rho}_r$是未知的，但在一些正则性条件下，可以通过样本协方差矩阵和样本相关矩阵相合地估计. 假定收益率是弱平稳的，且数据为$\{\boldsymbol{r}_t \mid t=1, \cdots, T\}$，则我们有如下估计：

$$\hat{\Sigma}_r \equiv [\hat{\sigma}_{ij,r}] = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{r}_t-\bar{\boldsymbol{r}})(\boldsymbol{r}_t-\bar{\boldsymbol{r}})', \quad \bar{\boldsymbol{r}}=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{r}_t \tag{8.37}$$

$$\hat{\boldsymbol{\rho}}_r = \hat{\boldsymbol{D}}^{-1}\hat{\Sigma}_r\hat{\boldsymbol{D}}^{-1} \tag{8.38}$$

其中$\hat{\boldsymbol{D}}=\mathrm{diag}\{\sqrt{\hat{\sigma}_{11,r}}, \cdots, \sqrt{\hat{\sigma}_{kk,r}}\}$是由$\boldsymbol{r}_t$的样本标准差构成的对角矩阵. 从而可以利用计算对称矩阵的特征值和特征向量的方法来进行主成分分析. 现在大多数统计包都能进行主成分分析.

例 8.7 考虑 IBM，惠普，Intel 公司，Merrill Lynch 与 Morgan Stanley Dean Witter 从 1990 年 1 月至 1999 年 12 月的月对数收益率. 此收益率以百分比表示，且包括红利. 数据集共有 120 个观测值. 图 8-10 给出了这五种月收益率序列的时间图. 如所料想，同一工业部门的公司收益率倾向于展现出相似的模式.

338 用$\boldsymbol{r}'=$(IBM，HWP，INTC，MER，MWD)表示这些收益率，其样本均值向量为(1.47，1.97，3.05，2.30，2.36)′，样本协方差矩阵和相关矩阵为

$$\hat{\Sigma}_r = \begin{bmatrix} 73.10 & & & & \\ 36.48 & 103.60 & & & \\ 27.08 & 48.86 & 113.96 & & \\ 16.06 & 37.59 & 27.96 & 105.56 & \\ 16.33 & 40.72 & 26.86 & 85.47 & 109.91 \end{bmatrix},$$

$$\hat{\boldsymbol{\rho}}_r = \begin{bmatrix} 1.00 & & & & \\ 0.42 & 1.00 & & & \\ 0.30 & 0.45 & 1.00 & & \\ 0.18 & 0.36 & 0.26 & 1.00 & \\ 0.18 & 0.38 & 0.24 & 0.79 & 1.00 \end{bmatrix}.$$

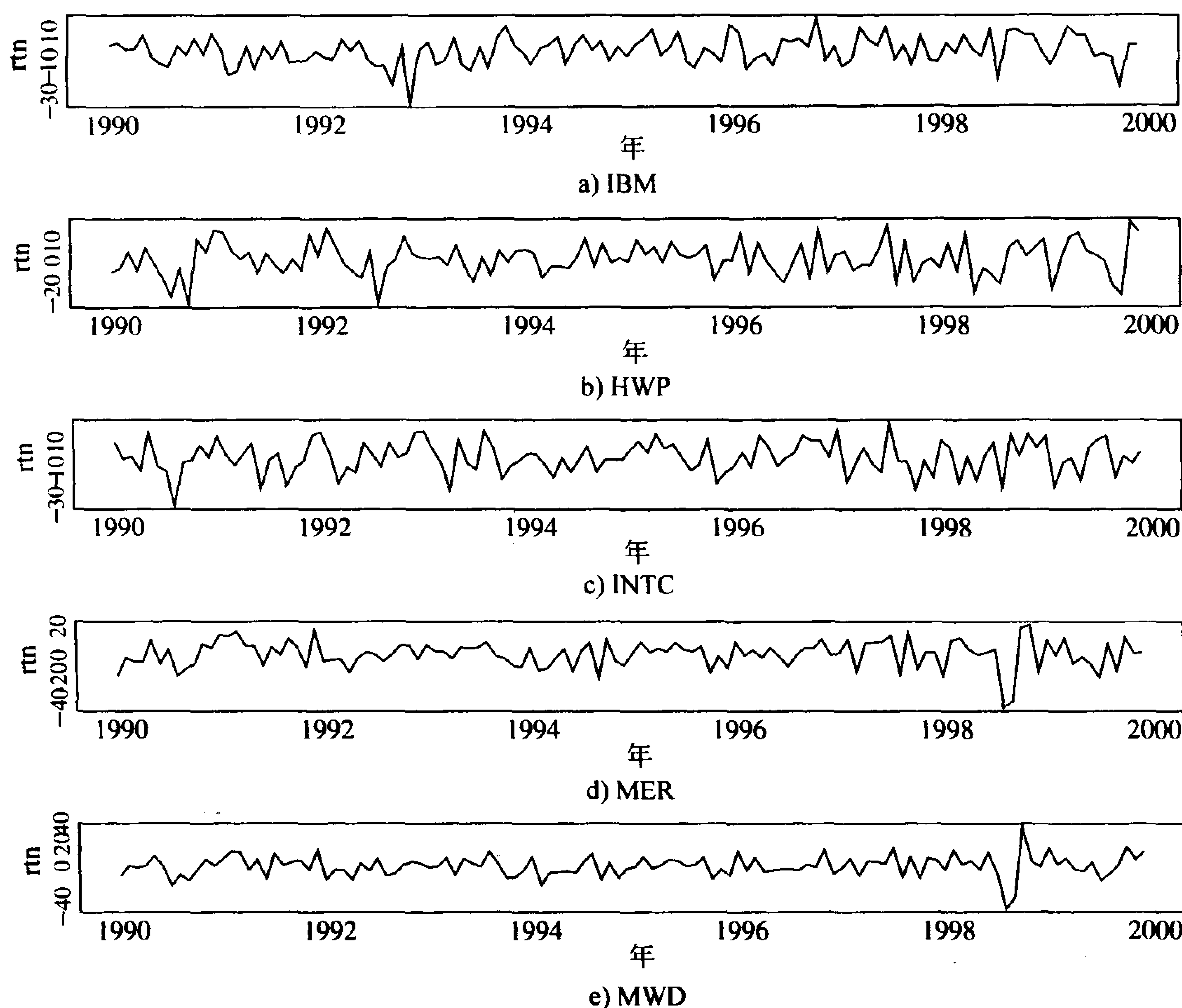


图 8-10 以百分比表示的 IBM，惠普，Intel，Merrill Lynch 和 Morgan Stanley Dean Witter 等公司从 1990 年 1 月到 1999 年 12 月的月对数收益率包括红利的时间图

表 8-8 给出了利用协方差矩阵和相关矩阵进行主成分分析的结果. 还给出了特征值、特征向量以及由主成分解释的变化比. 考虑相关矩阵，并用$\hat{\lambda}_i$ 和$\hat{\boldsymbol{e}}_i$ 来表示样本特征值与特征向量. 对前两个主成分，我们有 339

$$\hat{\lambda}_1 = 2.456, \quad \hat{\boldsymbol{e}}_1 = (0.342, 0.474, 0.387, 0.503, 0.505)',$$

$$\hat{\lambda}_2 = 1.145, \quad \hat{\boldsymbol{e}}_2 = (0.525, 0.314, 0.405, -0.481, -0.481)'.$$

这两个成分大约解释了数据全部变化的 72%，且它们具有有趣的解释. 第一个成分是股票收益率的一个大致为等权重的线性组合，这个成分可能代表股票市场的一般运动，从而是一个市场成分. 第二个成分代表两个工业部门（即技术和金融服务）的差. 它可能是一个工业成分. 利用 $\boldsymbol{r}$ 的协方差矩阵也可以发现主成分的类似解释.

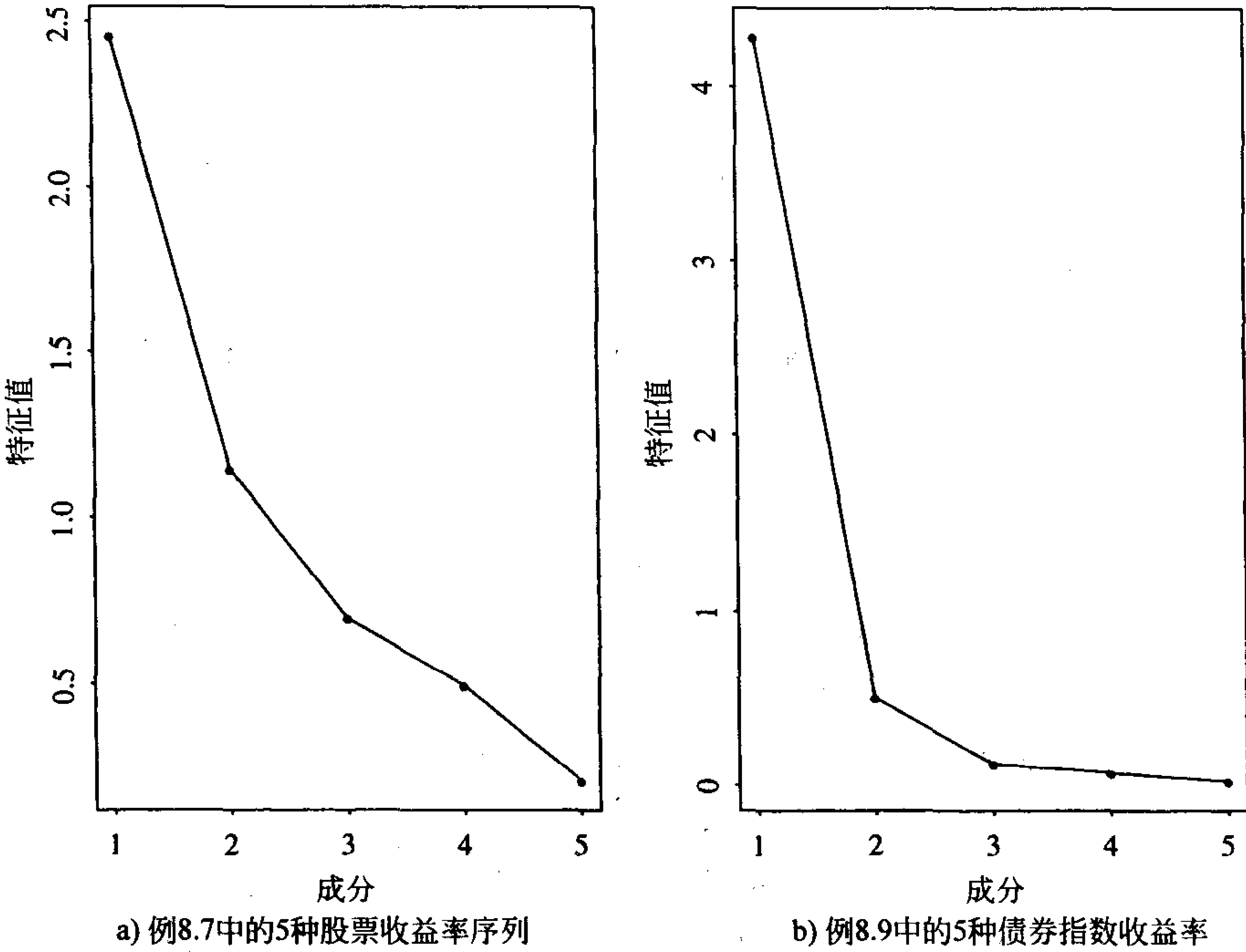
应用中需要的一个决定主成分数量的非正式但是很有用的方法是检查斜坡图(scree plot)，它是特征值按由大到小次序排列之后的时间图（即$\hat{\lambda}_i$ 对 i 的图）. 图 8-11a 给出了例 8.7 中的五种股票收益率的斜坡图. 通过在斜坡图中寻找转弯处，表明余下的特征值相对较小并大致看上去是相同的，可以选择一个恰当的主成分的个数. 对图 8-11 的两个图来说，两个主成分看起

表 8-8

	(a)利用样本协方差矩阵				
特征值	256.16	116.14	64.91	46.82	22.11
比例	0.506	0.229	0.128	0.093	0.044
累积	0.506	0.736	0.864	0.956	1.000
特征向量	0.246	0.327	0.586	−0.700	0.018
	0.461	0.360	0.428	0.687	−0.050
	0.409	0.585	−0.683	−0.153	0.033
	0.522	−0.452	−0.082	−0.115	−0.710
	0.536	−0.467	−0.036	−0.042	0.701
	(b)利用样本相关阵				
特征值	2.4563	1.1448	0.6986	0.4950	0.2053
比例	0.491	0.229	0.140	0.099	0.041
累积	0.491	0.720	0.860	0.959	1.000
特征向量	0.342	0.525	0.691	−0.362	−0.012
	0.474	0.314	−0.043	0.820	0.050
	0.387	0.405	−0.717	−0.414	−0.034
	0.503	−0.481	0.052	−0.147	0.701
	0.505	−0.481	0.071	−0.062	−0.711

注：对 IBM，惠普，Intel，Merrill Lynch 与 Morgan Stanley Dean Witter 从 1990 年 1 月至 1999 年 12 月的月对数收益率的主成分分析的结果. 特征向量以列向量形式给出.

340



a) 例8.7中的5种股票收益率序列

b) 例8.9中的5种债券指数收益率

图 8-11 两个 5 维资产收益率的斜坡图

来是合适的. 最后，除了对 $j>i$ 有 $\lambda_j=0$ 的情形，选择前 i 个主成分仅仅提供了数据总方差的一个近似. 如果一个很小的 i 可以提供一个好的近似，则这种简化是有价值的.

8.8 因子分析

多元分析中的一个主要困难是"维数的祸害". 特别地，当模型的阶或时间序列的维数增加时，参数模型的参数数量也经常陡增. 通常要寻找简化方法来克服维数祸害的困难. 从经验的观点，多元数据经常表现出一些相似的模式，这表明数据中有潜藏的共同结构的存在. 因子分析是文献中可以利用的简化方法之一，它的目的是识别几个因子，使得能够解释数据的协方差与相关矩阵中的绝大部分变化.

341

传统的因子分析假定数据没有序列相关性. 由于金融数据经常是以小于或等于一周的频率抽取的，所以经常违反这个假定. 然而，这个假定对低频数据的资产收益率(如股票或市场指数的月收益率)看上去是合理的. 如果违背了这个假定，则可以利用本书中讨论的参数模型消除数据的线性动态依赖，并对残差序列应用因子分析.

下面，我们讨论基于正交因子模型(orthogonal factor model)的因子分析. 考虑 k 维对数收益率 $\boldsymbol{r}=(r_1, \cdots, r_k)'$，并假定 $\boldsymbol{r}$ 的均值和协方差矩阵分别为 $\boldsymbol{\mu}$ 和 Σ. 对一个收益率序列，这个假定等价于要求 $\boldsymbol{r}$ 是弱平稳的. 因子模型假定 $\boldsymbol{r}$ 线性地依赖于少数不可观测的随机变量 $\boldsymbol{F}=(f_1, f_2, \cdots, f_m)'$ 与 k 维附加噪声 $\boldsymbol{\varepsilon}=(\varepsilon_1, \cdots, \varepsilon_k)'$. 这里 $m<k$，f_i 是公共因子，ε_i 是误差. 数学上，因子模型为

$$\begin{aligned}
r_1-\mu_1 &= l_{11}f_1+\cdots+l_{1m}f_m+\varepsilon_1\\
r_2-\mu_2 &= l_{21}f_1+\cdots+l_{2m}f_m+\varepsilon_2\\
\vdots &= \vdots\\
r_k-\mu_k &= l_{k1}f_1+\cdots+l_{km}f_m+\varepsilon_k
\end{aligned}$$

或等价地用矩阵记号

$$\boldsymbol{r}-\boldsymbol{\mu}=\boldsymbol{LF}+\boldsymbol{\varepsilon} \tag{8.39}$$

其中 $\boldsymbol{L}=[l_{ij}]_{k\times m}$ 是因子负荷矩阵，l_{ij} 是第 i 个变量对第 j 个因子的负荷，ε_i 是 r_i 的个性误差. 上述因子模型的一个关键特征是 m 个因子 f_i 和 k 个误差 ε_i 都是不可观测的. 因为这样，方程(8.39)不是一个多元线性回归模型，即使它看上去与多元线性回归模型相似.

方程(8.39)的因子模型是一个正交因子模型，如果它满足下面的假定：

1)$E(\boldsymbol{F})=\boldsymbol{0}$，$\mathrm{Cov}(\boldsymbol{F})=\boldsymbol{I}_m$，为 $m\times m$ 的单位矩阵；

2)$E(\boldsymbol{\varepsilon})=\boldsymbol{0}$，$\mathrm{Cov}(\boldsymbol{\varepsilon})=\boldsymbol{\psi}=\mathrm{diag}\{\psi_1, \cdots, \psi_k\}$(即 $\boldsymbol{\psi}$ 是 $k\times k$ 的对角矩阵)；

3)$\boldsymbol{F}$ 与 $\boldsymbol{\varepsilon}$ 是独立的，从而 $\mathrm{Cov}(\boldsymbol{F}, \boldsymbol{\varepsilon})=E(\boldsymbol{F\varepsilon}')=\boldsymbol{0}_{m\times k}$.

在上述假定下，很容易看出

$$\Sigma=\mathrm{Cov}(\boldsymbol{r})=E[(\boldsymbol{r}-\boldsymbol{\mu})(\boldsymbol{r}-\boldsymbol{\mu})']$$

$$=E[(\boldsymbol{LF}+\boldsymbol{\varepsilon})(\boldsymbol{LF}+\boldsymbol{\varepsilon})']$$
$$=\boldsymbol{L}E(\boldsymbol{FF}')\boldsymbol{L}'+E(\boldsymbol{\varepsilon F}')\boldsymbol{L}'+\boldsymbol{L}E(\boldsymbol{F\varepsilon}')+E(\boldsymbol{\varepsilon\varepsilon}')$$
342
$$=\boldsymbol{LL}'+\boldsymbol{\psi} \tag{8.40}$$

且

$$\mathrm{Cov}(\boldsymbol{r},\boldsymbol{F})=E[(\boldsymbol{r}-\boldsymbol{\mu})\boldsymbol{F}']=\boldsymbol{L}E(\boldsymbol{FF}')+E(\boldsymbol{\varepsilon F}')=\boldsymbol{L} \tag{8.41}$$

利用方程(8.40)和方程(8.41)，我们看出，对方程(8.39)中的正交因子模型

$$\mathrm{Var}(r_i)=l_{i1}^2+\cdots+l_{im}^2+\psi_i$$
$$\mathrm{Cov}(r_i,r_j)=l_{i1}l_{j1}+\cdots+l_{im}l_{jm}$$
$$\mathrm{Cov}(r_i,f_j)=l_{ij}.$$

由 m 个公共因子贡献的 r_i 的方差部分 $l_{i1}^2+\cdots+l_{im}^2$ 称为共性方差(Communality). r_i 方差的剩余部分 ψ_i 称为唯一性方差或个性方差. 令 $c_i^2=l_{i1}^2+\cdots+l_{im}^2$ 为共性方差，它是第 i 个变量对 m 个公共因子的负荷的平方和. 分量 r_i 的方差变为 $\mathrm{Var}(r_i)=c_i^2+\psi_i$.

实际中，并非每个协方差矩阵都具有正交因子表示. 换句话说，一个不具有任何正交因子表示的随机变量 $\boldsymbol{r}$ 是存在的. 而且，随机变量的正交因子并不唯一. 事实上，对任何满足 $\boldsymbol{PP}'=\boldsymbol{P}'\boldsymbol{P}=\boldsymbol{I}$ 的 $m\times m$ 正交矩阵 $\boldsymbol{P}$，令 $\boldsymbol{L}^*=\boldsymbol{LP}$，$\boldsymbol{F}^*=\boldsymbol{P}'\boldsymbol{F}$，则

$$\boldsymbol{r}-\boldsymbol{\mu}=\boldsymbol{LF}+\boldsymbol{\varepsilon}=\boldsymbol{LPP}'\boldsymbol{F}+\boldsymbol{\varepsilon}=\boldsymbol{L}^*\boldsymbol{F}^*+\boldsymbol{\varepsilon}.$$

另外 $E(\boldsymbol{F}^*)=\boldsymbol{0}$，$\mathrm{Cov}(\boldsymbol{F}^*)=\boldsymbol{P}'\mathrm{Cov}(\boldsymbol{F})\boldsymbol{P}=\boldsymbol{P}'\boldsymbol{P}=\boldsymbol{I}$. 这样，$\boldsymbol{L}^*$ 和 $\boldsymbol{F}^*$ 对 $\boldsymbol{r}$ 建立了另一个正交因子模型. 正交因子表示的这种不唯一性既是缺点，又是因子分析中的一个优点. 说它是缺点是因为它使得因子负荷的意义不确定了. 它也是一个优点，因为它允许我们进行旋转来寻找具有良好解释的公共因子. 因为 $\boldsymbol{P}$ 是一个正交矩阵，所以在 m 维空间中，变换 $\boldsymbol{F}^*=\boldsymbol{P}'\boldsymbol{F}$ 是一个旋转.

8.8.1 估计

方程(8.39)中的正交因子模型可以通过两种方法估计. 第一种方法利用前一节中的主成分分析. 这个方法不要求数据的正态性假定，也不要求预先指定公共因子的个数. 它对协方差矩阵和相关矩阵都是适用的. 但是同 PCA 中所提到的一样，这个解通常只是一个近似. 第二个估计方法是最大似然方法，它利用正态密度，并要求预先指定公共因子的个数.

主成分方法

343

再令 $(\hat{\lambda}_1,\ \hat{\boldsymbol{e}}_1)$，…，$(\hat{\lambda}_k,\ \hat{\boldsymbol{e}}_k)$ 是样本协方差矩阵 $\hat{\boldsymbol{\Sigma}}_r$ 的特征值-特征向量对，其中 $\hat{\lambda}_1\geqslant\hat{\lambda}_2\geqslant\cdots\geqslant\hat{\lambda}_k$. 令 $m<k$ 为公共因子的个数，则因子负荷矩阵由下式给出

$$\hat{\boldsymbol{L}}\equiv[\hat{l}_{ij}]=\left[\sqrt{\hat{\lambda}_1}\,\hat{\boldsymbol{e}}_1,\sqrt{\hat{\lambda}_2}\,\hat{\boldsymbol{e}}_2,\cdots,\sqrt{\hat{\lambda}_m}\,\hat{\boldsymbol{e}}_m\right] \tag{8.42}$$

估计的个性方差是矩阵 $\hat{\boldsymbol{\Sigma}}_r-\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}'$ 的主对角线上的元素. 即 $\hat{\boldsymbol{\psi}}=\mathrm{diag}\{\hat{\psi}_1,\ \cdots,\ \hat{\psi}_k\}$，其中 $\hat{\psi}_i=\hat{\sigma}_{ii,r}-\sum_{j=1}^{m}\hat{l}_{ij}^2$，$\hat{\sigma}_{ii,r}$ 是 $\hat{\boldsymbol{\Sigma}}_r$ 的第$(i,\ i)$个元素. 共性方差的估计为

$$\hat{c}_i^2 = \hat{l}_{i1}^2 + \cdots + \hat{l}_{im}^2$$

由近似产生的误差矩阵为

$$\hat{\Sigma}_r - (\hat{L}\hat{L}' + \hat{\psi}).$$

我们当然希望这个矩阵接近于0. 可以证明$\hat{\Sigma}_r - (\hat{L}\hat{L}' + \hat{\psi})$的元素的平方和小于或等于$\hat{\lambda}_{m+1}^2 + \cdots + \hat{\lambda}_k^2$. 因此，近似误差的上界为所忽略的特征值的平方和.

由方程(8.42)的解，基于主成分方法的因子的负荷估计并不随着公共因子 m 的增加而改变.

最大似然方法

如果公共因子 $\boldsymbol{F}$ 和个性因子 $\boldsymbol{\varepsilon}$ 是联合正态的，那么 $\boldsymbol{r}$ 是多元正态的，且均值为 $\boldsymbol{\mu}$，协方差矩阵为$\Sigma_r = \boldsymbol{LL}' + \boldsymbol{\psi}$. 在 $\boldsymbol{L}'\boldsymbol{\psi}^{-1}\boldsymbol{L} = \boldsymbol{\Delta}$(它是一个对角矩阵)的限制下，可以利用最大似然方法得到 $\boldsymbol{L}$ 和 $\boldsymbol{\psi}$ 的估计. 这里 $\boldsymbol{\mu}$ 是由样本均值估计的. 对这个方法的细节，读者可以参考 Johnson 和 Wichern(1998). 在利用最大似然方法时，公共因子的数量必须事先给定.

8.8.2 因子旋转

正如前面提到的，对任何 $m \times m$ 的正交矩阵 $\boldsymbol{P}$，

$$\boldsymbol{r} - \boldsymbol{\mu} = \boldsymbol{LF} + \boldsymbol{\varepsilon} = \boldsymbol{L}^* \boldsymbol{F}^* + \boldsymbol{\varepsilon},$$

其中 $\boldsymbol{L}^* = \boldsymbol{LP}$，$\boldsymbol{F}^* = \boldsymbol{P}'\boldsymbol{F}$. 另外，

$$\boldsymbol{LL}' + \boldsymbol{\psi} = \boldsymbol{LPP}'\boldsymbol{L}' + \boldsymbol{\psi} = \boldsymbol{L}^*(\boldsymbol{L}^*)' + \boldsymbol{\psi}.$$

这个结果说明：共性方差与个性方差在正交变换下保持不变. 那么寻找一个正交矩阵 $\boldsymbol{P}$ 来变换因子模型使得公共因子具有良好的性质就是合理的. 这样一个变换等价于将公共因子在 m 维空间中旋转. 事实上，有无限个可以利用的因子旋转. Kaiser(1958)提出了一个方差最大化准则(varimax criterion)来选择旋转，这在许多应用中都运作得很好. 记 $\boldsymbol{L}^* = [l_{ij}^*]$是旋转后的因子负荷矩阵，$c_i^2$ 表示第 i 个共性方差. 定义$\tilde{l}_{ij}^* = l_{ij}^* / c_i$ 为经过共性方差的(正)平方根的尺度变换之后的旋转系数. 方差最大化方法是选择对角矩阵 $\boldsymbol{P}$，使得下式最大化 [344]

$$V = \frac{1}{k}\sum_{j=1}^{m}\left[\sum_{i=1}^{k}(\tilde{l}_{ij}^*)^4 - \frac{1}{k}\left(\sum_{i=1}^{k}\tilde{l}_{ij}^{*2}\right)^2\right].$$

这个复杂的表示具有一个简单的解释. 最大化 V，对应于尽可能多地分散每个因子负荷的平方. 因此，此方法是为了在因子负荷的旋转矩阵的任何列中寻找大的但可以忽略的系数. 在实际应用中，使用因子旋转来帮助解释公共因子. 这在一些应用中可能有益，但在其他应用中未必有用.

8.8.3 应用

给定资产收益率的数据$\{\boldsymbol{r}_t\}$，因子分析使得我们能够找到解释收益率变化的公共因子. 因为因子分析假定数据没有序列相关性，所以在使用因子分析前应该检验这个假定的有效性. 为

此，我们可以使用多元混成统计量. 如果发现有序列相关性，则可以构造一个 VARMA 模型来消除数据中的动态相依性，并且对残差序列运用因子分析. 对许多收益率序列，线性模型残差的相关矩阵经常非常接近于原始数据的相关矩阵. 在这种情形下，动态依赖对因子分析的影响是可以忽略的.

例 8.8 再次考虑例 8.7 中使用的 IBM、HewLett-Parkard，Intel，Merrill Lynch 和 Morgan Stanley Dean Witter 的月对数股票收益率. 为了检验序列无关的假设，我们计算混成统计量得到 $Q_5(1)=34.28$，$Q_5(4)=114.30$，$Q_5(8)=216.78$. 与自由度为 25，100 和 200 的 χ^2 分布比较，这些检验统计量的 p 值分别是 0.102，0.156 和 0.198. 因此，即使在 10%的水平下都是不能拒绝序列无关的假设.

表 8-9 给出了基于相关矩阵运用主成分方法和最大似然方法的因子分析结果. 我们假定共同因子的数目是 2，根据例 8.7 中的主成分分析，这种取法是合理的. 从表中可见，因子分析揭示了几个有趣的发现：

345

- 由主成分方法识别的两个因子比最大似然方法识别的因子解释了更多的变化.
- 根据旋转后的因子负荷，两种方法对数据基本上识别了同样的两个公共因子. 金融股(Merrill Lynch 和 Morgan Stanley Dean Witter)对第一个因子的负荷很大，而技术股(IBM、Hewlett-Packard 和 Intel)对第二个因子的负荷很大. 这两个旋转后的因子联合区分了工业部门.
- 在这个特例中，方差最大化的旋转并没有将最大似然方法识别的两个因子改变很多. 然而由主成分方法识别的第一个非旋转因子被旋转去掉了. 根据方差最大化准则的思想，这并不奇怪.
- IBM 和 Intel 股票基于最大似然方法的收益率的具体方差相当大，暗示了这两只股票具有值得进一步研究的它们自己的特性.

表 8-9

(a)主成分法

变量	因子负荷的估计		旋转因子负荷		共性
	f_1	f_2	f_1^*	f_2^*	$1-\Psi_i$
IBM	0.536	0.561	0.011	0.776	0.602
HWP	0.744	0.335	0.317	0.752	0.665
INTC	0.607	0.433	0.151	0.730	0.556
MER	0.788	−0.515	0.928	0.158	0.887
MWD	0.791	−0.514	0.930	0.161	0.891
方差	2.456 3	1.144 8	1.850 2	1.750 9	3.601 1
比例	0.491	0.229	0.370	0.350	0.720

（续）

(b) 最大似然法

变量	因子负荷的估计		旋转因子负荷		共性
	f_1	f_2	f_1^*	f_2^*	$1-\Psi_i$
IBM	0.191	0.496	0.087	0.524	0.282
HWP	0.394	0.689	0.247	0.755	0.630
INTC	0.250	0.511	0.141	0.551	0.323
MER	0.800	0.072	0.769	0.232	0.645
MWD	0.994	−0.015	0.976	0.186	0.988
方差	1.881 3	0.986 6	1.632 4	1.235 5	2.867 9
比例	0.376	0.197	0.326	0.247	0.574

注：IBM、HewLett-Parkard，Intel，Merrill Lynch 和 Morgan Stanley Dean Witter 的月对数股票收益率的因子分析．收益率包含分红，数据是从 1990 年 1 月到 1999 年 12 月．分析基于样本交叉相关阵和两个公共因子．

346

例 8.9 在这个例子中，我们考虑期限为 30 年、20 年、10 年、5 年和 1 年的美国债券指数的月对数收益率．例 8.2 中描述过这个数据，但被转换成了对数收益率．总共有 696 个观测值．正如例 8.2 中显示的，数据中具有序列依赖性．然而，通过拟合一个 VARMA(2，1)模型来消除序列依赖几乎不对同步相关矩阵具有任何影响．事实上，拟合一个 VARMA(2，1)模型之前和之后的相关矩阵分别为

$$\hat{\boldsymbol{\rho}}_o=\begin{bmatrix}1.0 & & & & \\ 0.98 & 1.0 & & & \\ 0.92 & 0.91 & 1.0 & & \\ 0.85 & 0.86 & 0.90 & 1.0 & \\ 0.63 & 0.64 & 0.67 & 0.81 & 1.0\end{bmatrix},\quad \hat{\boldsymbol{\rho}}=\begin{bmatrix}1.0 & & & & \\ 0.98 & 1.0 & & & \\ 0.92 & 0.92 & 1.0 & & \\ 0.85 & 0.86 & 0.90 & 1.0 & \\ 0.66 & 0.67 & 0.71 & 0.84 & 1.0\end{bmatrix},$$

其中$\hat{\boldsymbol{\rho}}_o$是原始的对数收益率的相关矩阵．因此，我们直接对收益率序列应用因子分析．

表 8-10 中给出了数据因子分析的结果．对两种估计方法，前两个公共因子对数据总方差的解释都超过了 90%．事实上，高的共性方差说明对五种债券指数收益率而言，其个性方差都非常小．因为两种方法的结果接近，我们只讨论主成分方法．非旋转因子负荷说明：(a)所有五种收益序列对第一个因子的负荷粗略地相等．(b)对第二个因子的负荷与期限长短是正相关的．因此，第一个公共因子代表了一般的美国债券收益率，第二个因子体现了“期限”效应．而且，第二个因子负荷的和接近于 0，因此，这个公共因子也可以解释为长期债券与短期债券的比较．这里一个长期债券指的是期限为 10 年或更长的债券．对旋转后的因子，其负荷也是有趣的．对第一个旋转因子的负荷与期限成正比例，而对第二个因子的负荷与期限成反比例．

注释： 本节中的因子分析是利用 Minitab 计算机程序实现的．

表 8-10

(a)主成分法

变量	因子负荷的估计		旋转因子负荷		共性
	f_1	f_2	f_1^*	f_2^*	$1-\Psi_i$
30 年	0.952	0.253	0.927	0.333	0.970
20 年	0.954	0.240	0.922	0.345	0.968
10 年	0.956	0.140	0.866	0.429	0.934
5 年	0.955	−0.142	0.704	0.660	0.931
1 年	0.800	−0.585	0.325	0.936	0.982
方差	4.281 2	0.503 9	3.059 4	1.725 6	4.785 1
比例	0.856	0.101	0.612	0.345	0.957

(b)最大似然法

变量	因子负荷的估计		旋转因子负荷		共性
	f_1	f_2	f_1^*	f_2^*	$1-\Psi_i$
30 年	0.849	−0.513	0.895	0.430	0.985
20 年	0.857	−0.486	0.876	0.451	0.970
10 年	0.896	−0.303	0.744	0.584	0.895
5 年	1.000	0.000	0.547	0.837	1.000
1 年	0.813	0.123	0.342	0.747	0.675
方差	3.917 83	0.607 4	2.537 8	1.987 4	4.525 2
比例	0.784	0.121	0.508	0.397	0.905

注：期限为 30 年、20 年、10 年、5 年和 1 年的美国债券指数的月对数收益率的因子分析．数据是从 1942 年 1 月到 1999 年 12 月．分析基于样本交叉相关矩阵和两个公共因子．

附录 A 向量与矩阵的回顾

在这个附录中，我们主要回顾向量与矩阵的一些代数性质．这里没有给出证明，因为在关于矩阵的标准教科书（如 Graybill，1969）中可以找到．

一个 $m\times n$ 的实值矩阵是一个由 m 行 n 列的实数组成的数组．例如，

$$\mathbf{A}=\begin{bmatrix}2 & 5 & 8\\ -1 & 3 & 4\end{bmatrix},$$

是一个 2×3 的矩阵，该矩阵有 2 行 3 列．一般地，一个 $m\times n$ 矩阵可以写为

$$\mathbf{A}\equiv[a_{ij}]=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n}\\ \vdots & \vdots & & \vdots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{mn}\end{bmatrix} \tag{8.43}$$

正整数 m 和 n 为 $\boldsymbol{A}$ 的行数和列数，实数 a_{ij} 称为 $\boldsymbol{A}$ 的第(i,j)个元素，元素 a_{ii} 是矩阵的对角线元素.

347～348

一个 $m\times 1$ 矩阵形成一个 m 维列向量，一个 $1\times n$ 的矩阵是一个 n 维行向量. 文献中的向量一般指列向量. 如果 $m=n$，则矩阵是一个方阵. 如果对 $i\neq j$ 有$a_{ij}=0$，且 $m=n$，则矩阵 $\boldsymbol{A}$ 为一个对角矩阵. 如果对 $i\neq j$ 有 $a_{ij}=0$，而且对所有的 i，有 $a_{ii}=1$，则 $\boldsymbol{A}$ 是一个 $m\times m$ 的单位矩阵，通常用 $\boldsymbol{I}_m$ 表示，或当维数清楚时简单表示为 $\boldsymbol{I}$.

$n\times m$ 矩阵

$$\boldsymbol{A}'=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m-1,1} & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m-1,2} & a_{m2} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{m-1,n} & a_{mn} \end{bmatrix},$$

是矩阵 $\boldsymbol{A}$ 的转置. 例如，$\begin{bmatrix} 2 & -1 \\ 5 & 3 \\ 8 & 4 \end{bmatrix}$ 是 $\begin{bmatrix} 2 & 5 & 8 \\ -1 & 3 & 4 \end{bmatrix}$ 的转置. 我们利用记号 $\boldsymbol{A}'=[a'_{ij}]$ 来表示 $\boldsymbol{A}$ 的转置. 由定义得出 $a'_{ij}=a_{ji}$，且$(\boldsymbol{A}')'=\boldsymbol{A}$. 如果 $\boldsymbol{A}'=\boldsymbol{A}$，则 $\boldsymbol{A}$ 是一个对称矩阵.

基本运算

假设 $\boldsymbol{A}=[a_{ij}]_{m\times n}$，$\boldsymbol{C}=[c_{ij}]_{p\times q}$ 是两个矩阵，它们的维数在下标中给出. 令 b 是一个实数. 一些基本的矩阵运算定义如下：

- 加法：如果 $m=p$ 且 $n=q$，则 $\boldsymbol{A}+\boldsymbol{C}=[a_{ij}+c_{ij}]_{m\times n}$；
- 减法：如果 $m=p$，$n=q$，则 $\boldsymbol{A}-\boldsymbol{C}=[a_{ij}-c_{ij}]_{m\times n}$；
- 标量乘法：$b\boldsymbol{A}=[ba_{ij}]_{m\times n}$；
- 乘法：假设 $n=p$，则 $\boldsymbol{AC}=\left[\sum_{v=1}^{n}a_{iv}c_{vj}\right]_{m\times q}$.

当矩阵维数满足乘法的发生条件时，两个矩阵称为可相乘的(conformable). 下式是矩阵乘法的一个例子：

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & -4 \end{bmatrix}=\begin{bmatrix} 2\cdot 1-1\cdot 1 & 2\cdot 2+1\cdot 2 & 2\cdot 3-1\cdot 4 \\ 1\cdot 1-1\cdot 1 & 1\cdot 2+1\cdot 2 & 1\cdot 3-1\cdot 4 \end{bmatrix}$$

$$=\begin{bmatrix} 1 & 6 & 2 \\ 0 & 4 & -1 \end{bmatrix}$$

矩阵运算的重要法则包括：

(a) $(\boldsymbol{AC})'=\boldsymbol{C}'\boldsymbol{A}'$，

(b) 一般情况下，$\boldsymbol{AC}\neq\boldsymbol{CA}$.

349

逆、迹、特征值与特征向量

如果存在唯一的矩阵 $\boldsymbol{C}_{m\times m}$，满足 $\boldsymbol{AC}=\boldsymbol{CA}=\boldsymbol{I}_m$（$m\times m$ 的单位阵），则称方阵 $\boldsymbol{A}_{m\times m}$ 是非奇

异的或可逆的. 在这个情形，C 称为 A 的逆矩阵，记为 $C=A^{-1}$.

$A_{m\times m}$的迹是它的对角线元素的和$\left(\text{即 } tr(A)=\sum_{i=1}^{m}a_{ii}\right)$. 很容易看出，(a)$tr(A+C)=tr(A)+tr(C)$，(b)$tr(A)=tr(A')$，(c)假设两个矩阵是可相乘的，则$tr(AC)=tr(CA)$.

如果 $Ab=\lambda b$，则数 λ 与 $m\times 1$ 向量 b(可能是复值)就是矩阵 A 的一个右特征值与右特征向量对. 矩阵 A 有 m 个可能的特征值. 对实矩阵 A，复特征值是共轭出现的. 矩阵 A 是非奇异的当且仅当它的所有特征值都不为 0. 用$\{\lambda_i \mid i=1, \cdots, m\}$表示特征值，则我们有 $tr(A)=\sum_{i=1}^{m}\lambda_i$. 另外，矩阵 A 的行列式可以定义为 $|A|=\prod_{i=1}^{m}\lambda_i$. 对矩阵行列式的一般定义，可以参见关于矩阵的标准教材(如 Graybill，1969).

最后，矩阵 $A_{m\times n}$的秩是对称矩阵 AA'的非零特征值的数目. 对一个非奇异矩阵 A，还有$(A^{-1})'=(A')^{-1}$.

正定矩阵

设 $A(m\times m)$是方阵，如果(a)A 是对称的，(b)A 的所有特征值都是正数，则称 A 是一个正定矩阵. 另外一种定义：如果对任何非零的 m 维向量 b，都有$b'Ab>0$，则 A 是正定矩阵.

正定矩阵 A 的有用性质包括：(a)A 的所有特征值都是正实数，(b)矩阵 A 可以分解为 $A=P\Lambda P'$，其中 Λ 是一个包含 A 的所有特征值的对角矩阵，P 是一个包括 A 的 m 个右特征向量的 $m\times m$ 矩阵. 通常将特征值写为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_m$，相应的特征向量 e_1，…，e_m 满足 $Ae_i=\lambda_i e_i$ 且 $e_i'e_i=1$. 另外，如果特征值不同的话，这些特征向量彼此正交(即如果 $i\neq j$，则 $e_i'e_j=0$). 矩阵 P 是一个正交矩阵，这个分解称为矩阵 A 的谱分解. 例如，考虑简单的 2×2 矩阵

$$\Sigma=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix},$$

它是正定的. 简单的计算显示

350

$$\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix}=3\begin{bmatrix}1\\ 1\end{bmatrix},\quad \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix}=\begin{bmatrix}1\\ -1\end{bmatrix}.$$

因此，3 与 1 是Σ的特征值，标准化的特征向量分别为$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)'$和$\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)'$. 很容易证明如下谱分解成立

$$\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}.$$

对于对称矩阵 A，存在一个对角线上元素为 1 的下三角矩阵 L，以及对角矩阵 G，满足 $A=LGL'$；见 Strang(1980) 第 1 章. 如果 A 是正定的，则 G 的对角元素全是正的. 在这种情形，

$$A = L\sqrt{G}\sqrt{G}L' = L\sqrt{G}(L\sqrt{G})',$$

其中 $L\sqrt{G}$ 又是一个下三角矩阵，其平方根是逐个元素取的. 这个分解称为 A 的楚列斯基分解，它证明了正定矩阵 A 可以对角化为

$$L^{-1}A(L')^{-1} = L^{-1}A(L^{-1})' = G.$$

因为 L 是具有单位对角元素的下三角矩阵，所以 L^{-1} 也是具有单位对角元素的下三角矩阵. 再次考虑前面的 2×2 矩阵 Σ，容易验证

$$L = \begin{bmatrix} 1.0 & 0.0 \\ 0.5 & 1.0 \end{bmatrix}, \quad G = \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 1.5 \end{bmatrix}$$

满足 $\Sigma = LGL'$. 另外，

$$L^{-1} = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 1.0 \end{bmatrix}, \quad 且 \quad L^{-1}\Sigma(L^{-1})' = G.$$

拉直与克罗内克积

将 $m \times n$ 矩阵 A 按它的列写为 $A = [a_1, \cdots, a_n]$，定义拉直算子为 $\mathrm{vec}(A) = (a_1', a_2', \cdots, a_m')'$，它是一个 $mn \times 1$ 向量. 对两个矩阵 $A_{m\times n}$ 与 $C_{p\times q}$，A 与 C 的克罗内克积为

$$A \otimes C = \begin{bmatrix} a_{11}C & a_{12}C & \cdots & a_{1n}C \\ a_{21}C & a_{22}C & \cdots & a_{2n}C \\ \vdots & \vdots & & \vdots \\ a_{m1}C & a_{m2}C & \cdots & a_{mn}C \end{bmatrix}_{mp\times nq}$$

351

例如，假定

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 4 & -1 & 3 \\ -2 & 5 & 2 \end{bmatrix},$$

则 $\mathrm{vec}(A) = (2, -1, 1, 3)'$，$\mathrm{vec}(C) = (4, -2, -1, 5, 3, 2)'$，且

$$A \otimes C = \begin{bmatrix} 8 & -2 & 6 & 4 & -1 & 3 \\ -4 & 10 & 4 & -2 & 5 & 2 \\ -4 & 1 & -3 & 12 & -3 & 9 \\ 2 & -5 & -2 & -6 & 15 & 6 \end{bmatrix}.$$

假设维数是合适的，则对上述两个算子，我们有下面有用的性质：

1) 一般地，$A\otimes C \neq C\otimes A$；

2) $(A\otimes C)' = A'\otimes C'$；

3) $A\otimes(C+D) = A\otimes C + A\otimes D$；

4) $(A\otimes C)(F\otimes G) = (AF)\otimes(CG)$；

5) 如果 A 与 C 是可逆的，则 $(A\otimes C)^{-1} = A^{-1}\otimes C^{-1}$；

6) 对方阵 A 和 C，$tr(A\otimes C) = tr(A)tr(C)$；

7) $\mathrm{vec}(A+C) = \mathrm{vec}(A) + \mathrm{vec}(C)$；

8) $\text{vec}(\boldsymbol{ABC})=(\boldsymbol{C}'\otimes\boldsymbol{A})\text{vec}(\boldsymbol{B})$;

9) $tr(\boldsymbol{AC})=\text{vec}(\boldsymbol{C}')'\text{vec}(\boldsymbol{A})=\text{vec}(\boldsymbol{A}')'\text{vec}(\boldsymbol{C})$;

10)
$$\begin{aligned} tr(\boldsymbol{ABC}) &= \text{vec}(\boldsymbol{A}')'(\boldsymbol{C}'\otimes\boldsymbol{I})\text{vec}(\boldsymbol{B})=\text{vec}(\boldsymbol{A}')'(\boldsymbol{I}\otimes\boldsymbol{B})\text{vec}(\boldsymbol{C}) \\ &= \text{vec}(\boldsymbol{B}')'(\boldsymbol{A}'\otimes\boldsymbol{I})\text{vec}(\boldsymbol{C})=\text{vec}(\boldsymbol{B}')'(\boldsymbol{I}\otimes\boldsymbol{C})\text{vec}(\boldsymbol{A}) \\ &= \text{vec}(\boldsymbol{C}')'(\boldsymbol{B}'\otimes\boldsymbol{I})\text{vec}(\boldsymbol{A})=\text{vec}(\boldsymbol{C}')'(\boldsymbol{I}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{B}). \end{aligned}$$

在多元统计分析中，我们经常处理对称矩阵. 因此，可以很方便地将拉直算子推广为半拉直算子，它包含了主对角线或其以下的元素. 具体地，对一个对称方阵 $\boldsymbol{A}=[a_{ij}]_{k\times k}$，定义 $\text{vech}(\boldsymbol{A})=(\boldsymbol{a}_{1.}', \boldsymbol{a}_{2*}', \cdots, \boldsymbol{a}_{k*}')'$，其中 $\boldsymbol{a}_{1.}$ 是 $\boldsymbol{A}$ 的第一列，$\boldsymbol{a}_{i*}=(a_{ii}, a_{i+1,i}, \cdots, a_{ki})'$ 是 $(k-i+1)$ 维向量. $\text{vech}(\boldsymbol{A})$ 的维数是 $k(k+1)/2$. 例如，假设 $k=3$，则我们有 $\text{vech}(\boldsymbol{A})=(a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33})'$，它是一个 6 维向量.

352

附录 B 多元正态分布

一个 k 维随机向量 $\boldsymbol{x}=(x_1, \cdots, x_k)'$ 服从均值 $\boldsymbol{\mu}=(\mu_1, \cdots, \mu_k)'$，正定协方差矩阵 $\Sigma=[\sigma_{ij}]$ 的多元正态分布，如果它的概率分布密度函数(pdf)为

$$f(\boldsymbol{x}\mid\boldsymbol{\mu}, \Sigma)=\frac{1}{(2\pi)^{k/2}\mid\Sigma\mid^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right] \tag{8.44}$$

我们利用记号 $\boldsymbol{x}\sim N_k(\boldsymbol{\mu}, \Sigma)$ 表示 x 服从这样一个分布. 这个正态分布在多元统计分析中起着很重要的作用，它有几个很好的性质. 这里仅考虑与我们的研究有关的性质，对细节感兴趣的读者可以参考 Johnson 和 Wichern(1998).

为了深入了解多元正态分布，考虑二元情形(即取 $k=2$). 在这种情形，我们有

$$\Sigma=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}, \quad \Sigma^{-1}=\frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}\begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{bmatrix}.$$

利用相关系数 $\rho=\sigma_{12}/(\sigma_1\sigma_2)$，这里 $\sigma_i=\sqrt{\sigma_{ii}}$ 是 x_i 的标准差，我们有 $\sigma_{12}=\rho\sqrt{\sigma_{11}\sigma_{22}}$ 且 $\mid\Sigma\mid=\sigma_{11}\sigma_{22}(1-\rho^2)$，则 $\boldsymbol{x}$ 的概率密度函数变为

$$f(x_1, x_2\mid\boldsymbol{\mu}, \Sigma)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}[Q(\boldsymbol{x}, \boldsymbol{\mu}, \Sigma)]\right],$$

其中 $Q(\boldsymbol{x}, \boldsymbol{\mu}, \Sigma)=\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)$.

Johnson 和 Wichern(1998)第 4 章包含了这种概率密度函数的一些图形.

令 $\boldsymbol{c}=(c_1, \cdots, c_k)'$ 是一个非零的 k 维向量. 将随机向量分块为 $\boldsymbol{x}=(\boldsymbol{x}_1', \boldsymbol{x}_2')'$，其中 $\boldsymbol{x}_1=(x_1, \cdots, x_p)'$，$\boldsymbol{x}_2=(x_{p+1}, \cdots, x_k)'$，$1\leqslant p<k$. 相应地，将 $\boldsymbol{\mu}$ 和 Σ 也分块得到

$$\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}\sim N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right).$$

$\boldsymbol{x}$ 具有如下性质：

1) $\boldsymbol{c}'\boldsymbol{x}\sim N(\boldsymbol{c}'\boldsymbol{\mu}, \boldsymbol{c}'\Sigma\boldsymbol{c})$. 就是说，$\boldsymbol{x}$ 的任何非零线性组合是一元正态的. 这个性质反过来也成立. 具体地，如果对任何非零向量 $\boldsymbol{c}$，$\boldsymbol{c}'\boldsymbol{x}$ 是一元正态的，则 $\boldsymbol{x}$ 是多元正态的.

2) $\boldsymbol{x}_i$ 的边际分布是正态的. 事实上, 对 $i=1$ 和 2, $\boldsymbol{x}_i \sim N_{k_i}(\boldsymbol{\mu}_i, \Sigma_{ii})$, 其中 $k_1=p$, $k_2=k-p$.

353

3) $\Sigma_{12}=\mathbf{0}$ 当且仅当 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 独立.

4) 随机变量 $y=(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ 服从自由度为 m 的 χ^2 分布.

5) $\boldsymbol{x}_1$ 在给定 $\boldsymbol{x}_2=\boldsymbol{b}$ 下的条件分布也是正态的, 为

$$(\boldsymbol{x}_1 \mid \boldsymbol{x}_2=\boldsymbol{b}) \sim N_p[\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{b}-\boldsymbol{\mu}_2), \ \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}].$$

最后一个性质在许多科学领域都很有用. 例如, 它形成了正态假定下的时间序列预测与递推最小二乘估计的基础.

练习题

1. 考虑 Merck & Company, Johnson & Jonhson, GE, General Motors, Ford Motor Company 以及价值加权指数从 1960 年 1 月至 1999 年 12 月的以百分比表示的月对数股票收益率, 包括了红利支付, 见文件"m-mrk2vw. dat", 它共有 6 列, 次序同前面罗列的次序一致.
   (a) 计算数据的样本均值、样本协方差矩阵以及样本相关矩阵.
   (b) 检验原假设 H_0: $\boldsymbol{\rho}_1=\cdots=\boldsymbol{\rho}_6=\mathbf{0}$, 其中 $\boldsymbol{\rho}_i$ 为数据的延迟 i 的交叉相关矩阵. 基于 5% 显著水平推出结论.
   (c) 6 个收益率序列间有引导-延迟关系吗?
   (d) 利用样本协方差矩阵对数据进行主成分分析.
   (e) 利用样本相关矩阵对数据进行主成分分析.
   (f) 对数据进行因子分析. 识别公共因子的数目. 利用主成分和最大似然方法得到因子负荷的估计.
2. St Louis 联邦储备银行在网页上公布选择的利率及美国的金融数据, 网址是:

   http://www. stls. frb. org/fred/index. html

   考虑固定期限为 1 年和 10 年的国库券的月利率, 从 1953 年 4 月至 2000 年 10 月共 571 个观测值; 见文件"m-gsln10. dat". 利率是用百分比表示的.
   (a) 令 $c_t=r_t-r_{t-1}$ 为月利率 r_t 的变化序列. 对两个变化序列构造一个二元自回归模型. 讨论模型所蕴含的意义, 并将模型转换为结构形式.

354

   (b) 对两个变化序列建立一个二元滑动平均模型. 讨论这个模型所蕴含的意义, 并与前面的二元 AR 模型的结果比较.
   (c) 两个月利率序列是协整的吗?
3. 再次考虑固定期限为 1 年和 10 年的国库券的月利率, 从 1953 年 4 月至 2000 年 10 月共 571 个观测值. 考虑数据的对数序列, 并对序列建立一个 VARMA 模型. 讨论所得到模型蕴含的意义.
4. 再次考虑固定期限为 1 年和 10 年的国库券月利率, 从 1953 年 4 月至 2000 年 10 月. 这两个利率序列是门限协整的吗? 利用利差 $s_t=r_{10,t}-r_{1,t}$ 作为门限变量, 其中 r_{it} 指具有固定期限 i 年国库券利率. 如果它们是门限协整的, 对两个序列建立一个多元门限模型.

5. 二元 AR(4)模型 $\boldsymbol{x}_t-\boldsymbol{\Phi}_4\boldsymbol{x}_{t-4}=\phi_0+\boldsymbol{a}_t$ 是一个周期为 4 的特殊季节模型，$\{\boldsymbol{a}_t\}$是独立同分布的正态随机向量，均值为 0，协方差矩阵为Σ. 这种季节模型在研究公司的季度收入时可能有用.
   (a)假设 $\boldsymbol{x}_t$ 是弱平稳的，试推导 $\boldsymbol{x}_t$ 的均值向量与协方差矩阵.
   (b)推导 $\boldsymbol{x}_t$ 弱平稳性的充分必要条件.
   (c)证明对 $l>0$，$\boldsymbol{\Gamma}_l=\boldsymbol{\Phi}_4\boldsymbol{\Gamma}_{l-4}$，其中 $\boldsymbol{\Gamma}_l$ 是 $\boldsymbol{x}_t$ 的延迟 l 的自协方差矩阵.
6. 二元 MA(4)模型 $\boldsymbol{x}_t=\boldsymbol{a}_t-\boldsymbol{\Theta}_4\boldsymbol{a}_{t-4}$是周期为 4 的另一个季节模型，$\{\boldsymbol{a}_t\}$是独立同分布的正态随机向量，均值为 0，协方差矩阵为Σ. 试对 $l=0$，…，5，推导 $\boldsymbol{x}_t$ 的协方差矩阵 $\boldsymbol{\Gamma}_l$.

参考文献

Balke, N. S., and Fomby, T. B. (1997), "Threshold cointegration," *International Economic Review*, 38, 627–645.

Box, G. E. P., and Tiao, G. C. (1977), "A canonical analysis of multiple time series," *Biometrika*, 64, 355–366.

Brenner, R. J., and Kroner, K. F. (1995), "Arbitrage, cointegration, and testing the unbiasedness hypothesis in financial markets," *Journal of Financial and Quantitative Analysis*, 30 23–42.

Cochrane, J. H. (1988), "How big is the random walk in the GNP?" *Journal of Political Economy*, 96, 893–920.

Dwyer, Jr., G. P., Locke, P., and Yu, W. (1996), "Index arbitrage and nonlinear dynamics between the S&P 500 futures and cash," *Review of Financial Studies*, 9, 301–332.

Engle, R. F., and Granger, C. W. J. (1987), "Co-integration and error correction representation, estimation and testing," *Econometrica*, 55, 251–276.

355

Forbes, C. S., Kalb, G. R. J., and Kofman, P. (1999), "Bayesian arbitrage threshold analysis," *Journal of Business & Economic Statistics*, 17, 364–372.

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, Wiley: New York.

Graybill, F. A. (1969), *Introduction to Matrixes with Applications in Statistics*, Wadsworth: Belmont, California.

Hillmer, S. C., and Tiao, G. C. (1979), "Likelihood function of stationary multiple autoregressive moving average models," *Journal of the American Statistical Association*, 74, 652–660.

Hosking, J. R. M. (1980), "The multivariate portmanteau statistic," *Journal of the American Statistical Association*, 75, 602–608.

Hosking, J. R. M. (1981), "Lagrange-multiplier tests of multivariate time series models," *Journal of the Royal Statistical Society, Series B*, 43, 219–230.

Johansen, S. (1989), "Statistical analysis of co-integration vectors," *Journal of Economic Dynamics and Control*, 12,231–254.

Johnson, R. A. and Wichern, D. W. (1998), *Applied Multivariate Statistical Analysis*, 4th ed., Prentice Hall: Upper Saddle River, New Jersey.

Kaiser, H. F. (1958), "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, 23, 187–200.

Li, W. K., and McLeod, A. I. (1981), "Distribution of the residual autocorrelations in multivariate ARMA time series models," *Journal of the Royal Statistical Society, Series B*, 43, 231–239.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag: New York.

Reinsel, G. C. (1993), *Elements of Multivariate Time Series Analysis*, Springer-Verlag: New York.

Stock, J. H., and Watson, M. W. (1988), "Testing for common trends," *Journal of the American Statistical Association*, 83, 1097–1107.

Strang, G. (1980), *Linear Algebra and its Applications*, 2nd ed., Harcourt Brace Jovanovich: Chicago.

Tiao, G. C., and Box, G. E. P. (1981), "Modeling multiple time series with applications," *Journal of the American Statistical Association*, 76, 802–816.

Tiao, G. C., and Tsay, R. S. (1989), "Model specification in multivariate time series" (with discussions), *Journal of the Royal Statistical Society, Series B*, 51, 157–213.

Tiao, G. C., Tsay, R. S., and Wang, T. (1993), "Usefulness of linear transformations in multivariate time series analysis," *Empirical Economics*, 18, 567–593.

Tsay, R. S. (1991), "Two canonical forms for vector ARMA processes," *Statistica Sinica*, 1, 247–269.

Tsay, R. S. (1998), "Testing and modeling multivariate threshold models," *Journal of the American Statistical Association*, 93, 1188–1202.

Tsay, R. S., and Tiao, G. C. (1990), "Asymptotic properties of multivariate nonstationary processes with applications to autoregressions," *Annals of Statistics*, 18, 220–250.

356

第9章 多元波动率模型及其应用

本章中，我们将第3章中的一元波动率模型推广到多元情形，并讨论一些方法来简化多个资产收益率的波动率过程之间的动态关系. 多元波动率在金融中有许多重要的应用. 它们在组合选择与资产分配中非常重要，而且可以用来计算包括多种资产的金融头寸的风险值.

考虑多元收益率序列$\{\boldsymbol{r}_t\}$. 我们采取与一元情形相同的方法将序列改写为

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{a}_t$$

这里$\boldsymbol{\mu}_t = E(\boldsymbol{r}_t \mid \boldsymbol{F}_{t-1})$是$\boldsymbol{r}_t$在给定过去信息$\boldsymbol{F}_{t-1}$下的条件期望，$\boldsymbol{a}_t = (a_{1t}, \cdots, a_{kt})'$是序列在$t$时刻的抖动或新息. 假定$\boldsymbol{\mu}_t$过程服从第8章中多元时间序列模型的条件期望. 对大多数收益率序列，对$\boldsymbol{\mu}_t$采用一个简单的向量ARMA结构就足够了，即

$$\boldsymbol{\mu}_t = \phi_0 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{r}_{t-i} - \sum_{i=1}^{q} \boldsymbol{\Theta}_i \boldsymbol{a}_{t-i} \tag{9.1}$$

其中p，q为非负整数. 如有必要，可在上面的方程中加入解释变量. 我们称方程(9.1)为$\boldsymbol{r}_t$的均值方程.

在给定$\boldsymbol{F}_{t-1}$下$\boldsymbol{a}_t$的条件协方差矩阵是一个$k \times k$的正定矩阵Σ_t，定义为$\Sigma_t = \text{Cov}(\boldsymbol{a}_t \mid \boldsymbol{F}_{t-1})$. 给多元波动率建模关心的是$\Sigma_t$随时间的演变. 我们对$\Sigma_t$指定一个模型，称为收益率序列$\boldsymbol{r}_t$的波动率模型.

将一元波动率模型推广到多元情形有许多种方法，但是维数的繁琐很快成为应用中的一个主要阻碍，因为对一个k维收益率序列，Σ_t中有$k(k+1)/2$个量. 例如，对一个5维的收益率序列，Σ_t中有15个条件方差与条件协方差. 本章的目标是引进一些相对简单的多元波动率模型，它们很有用而且在实际应用中容易处理. 特别地，我们讨论了允许资产收益率间的相关系数是时变的一些模型. 时变相关性在金融中很有用. 例如，它们可以用来估计一个收益率序列的市场模型的时变β系数.

357

首先，在9.1节中，我们引进两种将波动率模型重新参数化的方法. 基于楚列斯基分解的重新参数化是非常有用的. 在9.2节中我们以GARCH模型为例，研究二元收益率的波动率模型. 在这个特例中，波动率模型可以是二元或三维的. 9.3节主要关注于更高维收益率的波动率模型. 9.4节强调维数降低的问题. 在9.5节中我们说明多元波动率模型的一些应用. 最后，在9.6节中，给出一个对波动率模型有用的多元学生-t分布.

9.1 重新参数化

对多元波动率建模时重要的一步就是利用Σ_t的对称性将它重新参数化. 我们考虑Σ_t的两个重新参数化.

9.1.1 相关系数的应用

对Σ_t的第一个重新参数化是利用$\boldsymbol{a}_t$的条件相关系数和方差．具体地，我们将Σ_t写为

$$\Sigma_t \equiv [\sigma_{ij,t}] = \boldsymbol{D}_t \boldsymbol{\rho}_t \boldsymbol{D}_t, \tag{9.2}$$

其中$\boldsymbol{\rho}_t$是$\boldsymbol{a}_t$的条件相关矩阵，$\boldsymbol{D}_t$是$k\times k$对角矩阵，包含了$\boldsymbol{a}_t$中元素的条件标准差（即$\boldsymbol{D}_t=\mathrm{diag}\{\sqrt{\sigma_{11,t}}, \cdots, \sqrt{\sigma_{kk,t}}\}$）．

因为$\boldsymbol{\rho}_t$是具有单位对角元的对称阵，所以Σ_t随时间演变是由条件方差$\sigma_{ii,t}$与$\boldsymbol{\rho}_t$的元素$\boldsymbol{\rho}_{ij,t}$控制的，其中$j<i$，$1\leqslant i\leqslant k$．因此，对$\boldsymbol{a}_t$的波动率建模，只要考虑a_{it}的条件方差和相关系数就足够了．定义$k(k+1)/2$维向量

$$\boldsymbol{\Xi}_t = (\sigma_{11,t}, \cdots, \sigma_{kk,t}, \boldsymbol{\varrho}'_t)', \tag{9.3}$$

其中$\boldsymbol{\varrho}_t$是由相关矩阵$\boldsymbol{\rho}_t$的列拉直得到的$k(k-1)/2$维向量，但是只利用了主对角下面的元素．具体写出来，对一个k维收益率序列，

$$\boldsymbol{\varrho}_t = (\rho_{21,t}, \cdots, \rho_{k1,t} \mid \rho_{32,t}, \cdots, \rho_{k2,t} \mid \cdots \mid \rho_{k,k-1,t})'$$

358

例如，对$k=2$，我们有$\boldsymbol{\varrho}_t=\rho_{21,t}$且

$$\boldsymbol{\Xi}_t = (\sigma_{11,t}, \sigma_{22,t}, \rho_{21,t})', \tag{9.4}$$

它是一个3维的向量．对$k=3$，我们有$\boldsymbol{\varrho}_t=(\rho_{21,t}, \rho_{31,t}, \rho_{32,t})'$且

$$\boldsymbol{\Xi}_t = (\sigma_{11,t}, \sigma_{22,t}, \sigma_{33,t}, \rho_{21,t}, \rho_{31,t}, \rho_{32,t})', \tag{9.5}$$

它是一个6维的向量．

如果$\boldsymbol{a}_t$是二元正态随机变量，则$\boldsymbol{\Xi}_t$由方程(9.4)给出，而且$\boldsymbol{a}_t$在给定$\boldsymbol{F}_{t-1}$下的条件密度函数为

$$f(a_{1t}, a_{2t} \mid \boldsymbol{\Xi}_t) = \frac{1}{2\pi\sqrt{\sigma_{11,t}\sigma_{22,t}(1-\rho_{21,t}^2)}} \exp\left[-\frac{Q(a_{1t}, a_{2t}, \boldsymbol{\Xi}_t)}{2(1-\rho_{21,t}^2)}\right],$$

其中$Q(a_{1t}, a_{2t}, \boldsymbol{\Xi}_t)=\dfrac{a_{1t}^2}{\sigma_{11,t}}+\dfrac{a_{2t}^2}{\sigma_{22,t}}-\dfrac{2\rho_{21,t}a_{1t}a_{2t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}}$．与最大似然估计有关的$a_t$的对数概率密度函数为

$$\ell(a_{1t}, a_{2t}, \boldsymbol{\Xi}_t) = -\frac{1}{2}\left\{\ln[\sigma_{11,t}\sigma_{22,t}(1-\rho_{21,t}^2)] + \frac{1}{1-\rho_{21,t}^2}\left(\frac{a_{1t}^2}{\sigma_{11,t}}+\frac{a_{2t}^2}{\sigma_{22,t}}-\frac{2\rho_{21,t}a_{1t}a_{2t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}}\right)\right\}. \tag{9.6}$$

这个重新参数化是很有用的，因为它直接对协方差和相关阵建模．然而，这个方法也有几个缺点．第一，当$k\geqslant 3$时，似然函数变得非常复杂．第二，此方法在估计中要求带限制的最大化，以保证Σ_t的正定性．当k很大时，这个限制变得很复杂．

9.1.2 楚列斯基分解

Σ_t的第二个重新参数化是利用楚列斯基分解；见第8章的附录A．这个方法在估计中有很多优点，因为它不会为了Σ_t的正定性加任何的参数限制；见Pourahmadi(1999)．另外，这个重新参数化是一个正交变换，所以似然函数极其简单．下面给出变换的细节．

359

因为 Σ_t 是正定的，所以存在一个具有单位对角元素的下三角矩阵 $\boldsymbol{L}_t$ 和具有正对角元素的对角矩阵 $\boldsymbol{G}_t$，满足

$$\sum\nolimits_t = \boldsymbol{L}_t \boldsymbol{G}_t \boldsymbol{L}'_t. \tag{9.7}$$

这就是著名的 Σ_t 的楚列斯基分解．此分解的一个良好性质是 $\boldsymbol{L}_t$ 对角线以下的元素与 $\boldsymbol{G}_t$ 的对角元素都有良好的解释．我们通过仔细研究二元和三维的情形来说明这个分解．对二元情形，我们有

$$\Sigma_t = \begin{bmatrix} \sigma_{11,t} & \sigma_{21,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix}, \quad \boldsymbol{L}_t = \begin{bmatrix} 1 & 0 \\ q_{21,t} & 1 \end{bmatrix}, \quad \boldsymbol{G}_t = \begin{bmatrix} g_{11,t} & 0 \\ 0 & g_{22,t} \end{bmatrix},$$

其中对 $i=1$ 和 2，有 $g_{ii,t}>0$．利用方程(9.7)，我们有

$$\Sigma_t = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} g_{11,t} & q_{21,t} g_{11,t} \\ q_{21,t} g_{11,t} & g_{22,t} + q_{21,t}^2 g_{11,t} \end{bmatrix}.$$

令上述矩阵方程的元素相等，我们得到

$$\sigma_{11,t} = g_{11,t}, \quad \sigma_{21,t} = q_{21,t} g_{11,t}, \quad \sigma_{22,t} = g_{22,t} + q_{21,t}^2 g_{11,t}. \tag{9.8}$$

解这个方程，我们有

$$g_{11,t} = \sigma_{11,t}, \quad q_{21,t} = \frac{\sigma_{21,t}}{\sigma_{11,t}}, \quad g_{22,t} = \sigma_{22,t} - \frac{\sigma_{21,t}^2}{\sigma_{11,t}}. \tag{9.9}$$

然而，考虑简单的条件线性回归

$$a_{2t} = \beta a_{1t} + b_{2t}, \tag{9.10}$$

其中 b_{2t} 表示误差项．由著名的最小二乘理论，我们有 $\beta = \dfrac{\mathrm{Cov}(a_{1t},\ a_{2t})}{\mathrm{Var}(a_{1t})} = \dfrac{\sigma_{21,t}}{\sigma_{11,t}}$，$\mathrm{Var}(b_{2t}) = \mathrm{Var}(a_{2t}) - \beta^2 \mathrm{Var}(a_{1t}) = \sigma_{22,t} - \dfrac{\sigma_{21,t}^2}{\sigma_{11,t}}$．而且，误差项 b_{2t} 与回归项 a_{1t} 是无关的．因此，利用方程(9.9)，我们有

360

$$g_{11,t} = \sigma_{11,t}, \quad q_{21,t} = \beta, \quad g_{22,t} = \mathrm{Var}(b_{2t}), \quad b_{2t} \perp a_{1t},$$

其中 $\perp$ 表示不相关．总之，2×2 矩阵 Σ_t 的楚列斯基分解相当于从 $\boldsymbol{a}_t$ 到 $\boldsymbol{b}_t = (b_{1t},\ b_{2t})'$ 进行的一个正交变换，满足

$$b_{1t} = a_{1t}, \quad b_{2t} = a_{2t} - q_{21,t} a_{1t},$$

其中 $q_{21,t}=\beta$ 是通过式(9.10)中的线性回归得到的，$\mathrm{Cov}(\boldsymbol{b}_t)$ 是对角元素为 $g_{ii,t}$ 的对角阵．变换的量 $q_{21,t}$ 与 $g_{ii,t}$ 可以做如下解释：

1) $\boldsymbol{G}_t$ 的第一个对角元恰是 a_{1t} 的方差．

2) $\boldsymbol{G}_t$ 的第二个对角元是方程(9.10)中简单线性回归的残差的方差．

3) 下三角矩阵 $\boldsymbol{L}_t$ 的元素 $q_{21,t}$ 是方程(9.10)中的回归系数 β．

对高维情形，以上性质仍然成立．例如，考虑三维情形

$$\boldsymbol{L}_t = \begin{bmatrix} 1 & 0 & 0 \\ q_{21,t} & 1 & 0 \\ q_{31,t} & q_{32,t} & 1 \end{bmatrix}, \quad \boldsymbol{G}_t = \begin{bmatrix} g_{11,t} & 0 & 0 \\ 0 & g_{22,t} & 0 \\ 0 & 0 & g_{33,t} \end{bmatrix}.$$

由方程(9.7)中的分解，我们有

$$\begin{bmatrix}\sigma_{11,t} & \sigma_{21,t} & \sigma_{31,t}\\ \sigma_{21,t} & \sigma_{22,t} & \sigma_{32,t}\\ \sigma_{31,t} & \sigma_{32,t} & \sigma_{33,t}\end{bmatrix}=\begin{bmatrix}g_{11,t} & q_{21,t}g_{11,t} & q_{31,t}g_{11,t}\\ q_{21,t}g_{22,t} & q_{21,t}^2g_{11,t}+g_{22,t} & q_{31,t}q_{21,t}g_{11,t}+q_{32,t}g_{22,t}\\ q_{31,t}g_{11,t} & q_{31,t}q_{21,t}g_{11,t}+q_{32,t}g_{22,t} & q_{31,t}^2g_{11,t}+q_{32,t}^2g_{22,t}+g_{33,t}\end{bmatrix}.$$

令上述矩阵方程中的元素相符，得到

$$\sigma_{11,t}=g_{11,t},\quad \sigma_{21,t}=q_{21,t}g_{11,t},\quad \sigma_{22,t}=q_{21,t}^2g_{11,t}+g_{22,t},\quad \sigma_{31,t}=q_{31,t}g_{11,t},$$

$$\sigma_{32,t}=q_{31,t}q_{21,t}g_{11,t}+q_{32,t}g_{22,t},\quad \sigma_{33,t}=q_{31,t}^2g_{11,t}+q_{32,t}^2g_{22,t}+g_{33,t}$$

或等价地，

$$g_{11,t}=\sigma_{11,t},\quad q_{21,t}=\frac{\sigma_{21,t}}{\sigma_{11,t}},\quad g_{22,t}=\sigma_{22,t}-q_{21,t}^2g_{11,t},$$

$$q_{31,t}=\frac{\sigma_{31,t}}{\sigma_{11,t}},\quad q_{32,t}=\frac{1}{g_{22,t}}\left(\sigma_{32,t}-\frac{\sigma_{31,t}\sigma_{21,t}}{\sigma_{11,t}}\right),\quad g_{33,t}=\sigma_{33,t}-q_{31,t}^2g_{11,t}-q_{32,t}^2g_{22,t}.$$

361

这些量看起来很复杂，但是它们只不过是下述正交变换的系数和残差方差

$$b_{1t}=a_{1t},\quad b_{2t}=a_{2t}-\beta_{21}b_{1t},\quad b_{3t}=a_{3t}-\beta_{31}b_{1t}-\beta_{32}b_{2t},$$

其中 β_{ij} 是下述最小二乘回归的系数.

$$a_{2t}=\beta_{21}b_{1t}+b_{2t},$$
$$a_{3t}=\beta_{31}b_{1t}+\beta_{32}b_{2t}+b_{3t}.$$

换句话说，我们有

$$q_{ij,t}=\beta_{ij},g_{ii,t}=\mathrm{Var}(b_{it})\text{ 且对 }i\neq j,\text{有 }b_{it}\perp b_{jt}.$$

根据前面的讨论，利用楚列斯基分解相当于做一个从 $\boldsymbol{a}_t$ 到 $\boldsymbol{b}_t$ 的正交变换，其中，$b_{1t}=a_{1t}$ 且对于 $1<i\leqslant k$，b_{it} 是通过最小二乘回归递推定义的：

$$b_{it}=q_{i1,t}b_{1t}+q_{i2,t}b_{2t}+\cdots+q_{i(i-1),t}b_{(i-1)t}+b_{it},\tag{9.11}$$

其中对于 $1\leqslant j<i$，$q_{ij,t}$ 是下三角矩阵 L_t 的第(i,j)个元素. 我们可以将这个变换写为

$$\boldsymbol{b}_t=\boldsymbol{L}_t^{-1}\boldsymbol{a}_t,\quad \text{或 }\boldsymbol{a}_t=\boldsymbol{L}_t\boldsymbol{b}_t,\tag{9.12}$$

其中，如前所述，$\boldsymbol{L}_t^{-1}$ 也是一个具有单位对角元素的下三角阵. $\boldsymbol{b}_t$ 的协方差矩阵是楚列斯基分解中的对角阵 $\boldsymbol{G}_t$，因为 $\mathrm{Cov}(\boldsymbol{b}_t)=\boldsymbol{L}_t^{-1}\Sigma_t(\boldsymbol{L}_t^{-1})'=\boldsymbol{G}_t$. 在这样一个变换下，与波动率建模相关的参数向量变为

$$\boldsymbol{\Xi}_t=(g_{11,t},\cdots,g_{kk,t},q_{21,t},q_{31,t},q_{32,t},\cdots,q_{k1,t},\cdots,q_{k(k-1),t})',\tag{9.13}$$

它也是一个 $k(k+1)/2$ 维的向量.

前面的正交变换也大大简化了数据的似然函数. 利用 $|\boldsymbol{L}_t|=1$ 的事实，我们有

$$|\Sigma_t|=|\boldsymbol{L}_t\boldsymbol{G}_t\boldsymbol{L}_t'|=|\boldsymbol{G}_t|=\prod_{i=1}^{k}g_{ii,t}.\tag{9.14}$$

如果 $\boldsymbol{a}_t$ 在给定过去信息下的条件分布是多元正态 $N(\boldsymbol{0},\Sigma_t)$，则变换后的序列 $\boldsymbol{b}_t$ 的条件分布 362

也是多元正态 $N(\mathbf{0}, \boldsymbol{G}_t)$，且数据的对数似然函数变得极其简单．事实上，我们有 $\boldsymbol{a}_t$ 的对数概率密度为

$$\ell(\boldsymbol{a}_t, \sum\nolimits_t) = \ell(\boldsymbol{b}_t, \boldsymbol{\Xi}_t) = -\frac{1}{2}\sum_{i=1}^{k}\left[\ln(g_{ii,t}) + \frac{b_{it}^2}{g_{ii,t}}\right], \tag{9.15}$$

这里为了简便，忽略了常数项，且 $g_{ii,t}$ 为 b_{it} 的方差．

利用楚列斯基分解，将 Σ_t 重新参数化有几个优点．第一，由方程(9.14)，如果对所有的 i，有 $g_{ii,t}>0$，则 Σ_t 是正定的．因此，Σ_t 的正定限制可以通过对 $\ln(g_{ii,t})$ 而不是 $g_{ii,t}$ 建模很容易地得到．第二，方程(9.13)中参数向量 $\boldsymbol{\Xi}_t$ 的元素有良好的解释．它们是使得抖动与收益率正交化的多元线性回归的系数与残差方差．第三，a_{1t} 与 a_{2t} 间的相关系数为

$$\rho_{21,t} = \frac{\sigma_{21,t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}} = q_{21,t} \times \frac{\sqrt{\sigma_{11,t}}}{\sqrt{\sigma_{22,t}}},$$

如果 $q_{21,t} \neq \rho\sqrt{\sigma_{22,t}}/\sqrt{\sigma_{11,t}}$，其中 ρ 是常数，则 $\rho_{21,t}$ 是时变的．例如，如果 $q_{21,t}=c\neq 0$，则 $\rho_{21,t}=c\sqrt{\sigma_{11,t}}/\sqrt{\sigma_{22,t}}$，假定方差比 $\sigma_{11,t}/\sigma_{22,t}$ 不是常数，则它仍然是时变的．若 r_t 的维数大于 2，时变性也适用于其他相关系数，并且时变性是 Σ_t 重新参数化的两种方法间的主要区别．

利用方程(9.11)及变换后的抖动 b_{it} 之间的正交性，我们得到

$$\sigma_{ii,t} = \mathrm{Var}(a_{it} \mid F_{t-1}) = \sum_{v=1}^{i} q_{iv,t}^2 g_{vv,t}, \qquad i = 1,\cdots,k,$$

$$\sigma_{ij,t} = \mathrm{Cov}(a_{it}, a_{jt} \mid F_{t-1}) = \sum_{v=1}^{j} q_{iv,t} q_{jv,t} g_{vv,t}, \qquad j < i, i = 2,\cdots,k,$$

其中对 $v=1, \cdots, k$，$q_{vv,t}=1$．这些方程给出了 Σ_t 在楚列斯基分解下的参数化．

9.2 二元收益率的 GARCH 模型

因为可以用同样的方法将许多一元的波动率模型推广到多元情形，所以我们将讨论集中于多元 GARCH 模型．其他的多元波动率模型也可以用．

对一个 k 维收益率序列 $\boldsymbol{r}_t$，多元 GARCH 模型利用“精确方程”来描述 $k(k+1)/2$ 维向量 $\boldsymbol{\Xi}_t$ 随时间的演变．精确方程的意思是这个方程不包含任何的随机抖动．然而，精确方程即使是在最简单的情形 $k=2$ 时也可能变得很复杂，此时 $\boldsymbol{\Xi}_t$ 是三维的．为了保持模型的简单性，通常对方程加一些限制．

363

9.2.1 常相关模型

为了保持波动率方程数量较少，Bollerslev(1990)考虑了相关系数 $\rho_{21,t}=\rho_{21}$ 不是时变的这种特殊情况，其中 $|\rho_{21}|<1$．在这个假定下，ρ_{21} 是一个常参数，从而对 $\boldsymbol{\Xi}_t^* = (\sigma_{11,t}, \sigma_{22,t})'$ 的波动率模型包含两个参数．$\boldsymbol{\Xi}_t^*$ 的 GARCH(1，1)模型变为

$$\boldsymbol{\Xi}_t^* = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 \boldsymbol{a}_{t-1}^2 + \boldsymbol{\beta}_1 \boldsymbol{\Xi}_{t-1}^*, \tag{9.16}$$

其中 $\boldsymbol{a}_{t-1}^2 = (a_{1,t-1}^2, a_{2,t-1}^2)$，$\boldsymbol{\alpha}_0$ 是一个二维正向量，$\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\beta}_1$ 是 2×2 非负定矩阵．更具体地，

此模型可以更详细地表示为

$$\begin{bmatrix}\sigma_{11,t}\\ \sigma_{22,t}\end{bmatrix}=\begin{bmatrix}\alpha_{10}\\ \alpha_{20}\end{bmatrix}+\begin{bmatrix}\alpha_{11} & \alpha_{12}\\ \alpha_{21} & \alpha_{22}\end{bmatrix}\begin{bmatrix}a_{1,t-1}^2\\ a_{2,t-1}^2\end{bmatrix}+\begin{bmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1}\\ \sigma_{22,t-1}\end{bmatrix}, \tag{9.17}$$

其中对 $i=1$ 和 2，有 $\alpha_{i0}>0$. 定义 $\boldsymbol{\eta}_t=\boldsymbol{a}_t^2-\boldsymbol{\Xi}_t^*$，我们可将前面的模型改写为

$$\boldsymbol{a}_t^2=\boldsymbol{\alpha}_0+(\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1)\boldsymbol{a}_{t-1}^2+\boldsymbol{\eta}_t-\boldsymbol{\beta}_1\boldsymbol{\eta}_{t-1}$$

这是对 $\boldsymbol{a}_t^2$ 过程的一个二元 ARMA(1，1)模型. 这个结果是第 3 章中一元 GARCH(1，1)模型的直接推广. 因此，模型(9.17)的一些性质可以从第 8 章中的二元 ARMA(1，1)模型的性质中很容易地得到. 尤其是，我们有下面的结果：

1)如果$\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1$ 所有的特征值都是正的，并小于 1，则 $\boldsymbol{a}_t^2$ 的二元 ARMA(1，1)模型是弱平稳的，从而 $E(\boldsymbol{a}_t^2)$存在. 这蕴含了收益率的扰动过程 $\boldsymbol{a}_t$ 有正定的无条件协方差矩阵. $\boldsymbol{a}_t$ 元素的无条件方差是$(\sigma_1^2，\sigma_2^2)'=(\boldsymbol{I}-\boldsymbol{\alpha}_1-\boldsymbol{\beta}_1)^{-1}\phi_0$，且 a_{1t}与 a_{2t}间的无条件协方差是 $\rho_{21}\sigma_1\sigma_2$.

2)如果 $\alpha_{12}=\beta_{12}=0$，则 a_{1t}的波动率不依赖于 a_{2t}过去的波动率. 类似地，如果 $\alpha_{21}=\beta_{21}=0$，则 a_{2t}的波动率不依赖于 a_{1t}过去的波动率.

3)如果$\boldsymbol{\alpha}_1$ 和$\boldsymbol{\beta}_1$ 是对角的，则此模型简化为两个一元 GARCH(1，1)模型. 在这种情形，这两个波动率过程不是动态相关的.

4)可以利用向量 ARMA(1，1)模型的预测方法类似地得到该模型的波动率预测，见第 3 章中的一元情形. 在预测原点 h 的 1 步向前波动率预测为

364

$$\boldsymbol{\Xi}_h^*(1)=\boldsymbol{\alpha}_0+\boldsymbol{\alpha}_1\boldsymbol{a}_h^2+\boldsymbol{\beta}_1\boldsymbol{\Xi}_h^*$$

对 ℓ 步向前预测，我们有

$$\boldsymbol{\Xi}_h^*(\ell)=\boldsymbol{\alpha}_0+(\boldsymbol{\alpha}_1+\boldsymbol{\beta}_1)\boldsymbol{\Xi}_h^*(\ell-1),\qquad \ell>1.$$

这些预测是对 a_{it} 的边际波动率的预测. a_{1t}与 a_{2t}之间的向前 ℓ 步协方差预测为$\hat{\rho}_{21}[\sigma_{11,h}(\ell)\sigma_{22,h}(\ell)]^{0.5}$，其中$\hat{\rho}_{21}$是 ρ_{21}的估计，$\sigma_{ii,h}(\ell)$是$\boldsymbol{\Xi}_h^*(\ell)$的元素.

例 9.1 考虑香港和日本的股票市场指数从 1996 年 1 月 1 日至 1997 年 10 月 16 日的日对数收益率，共 469 个观测值. 指数是以美元度量的，收益率是百分比表示的. 选择这个样本段是为了避免亚洲金融危机的影响，该危机在 1997 年 10 月 17 日冲击了香港市场. 数据来自于数据流(Datastream). 图 9-1 给出了两个指数收益率的时间图. 令 r_{1t}为香港股票市场的指数收益率，r_{2t}为日本股票市场的指数收益率. 如果采用一元 GARCH 模型，则我们得到这个模型

$$r_{1t}=0.137r_{1,t-1}+a_{1t},\quad a_{1t}=\sigma_{1t}\varepsilon_{1t}$$
$$\sigma_{1t}^2=0.164+0.142a_{1,t-1}^2+0.765\sigma_{1,t-1}^2 \tag{9.18}$$
$$r_{2t}=a_{2t},\quad a_{2t}=\sigma_{2t}\varepsilon_{2t},$$
$$\sigma_{2t}^2=0.085+0.128a_{2,t-1}^2+0.807\sigma_{2,t-1}^2 \tag{9.19}$$

其中除了 r_{1t}序列的 AR(1)系数(其 p 值为 0.029)以外的所有参数的估计都是高度显著的. 上述两个模型的标准化残差与残差平方序列的 L-B 统计量并没有表明模型的任何不合适性. 图 9-2 给出了前面两个一元 GARCH(1，1)模型的波动率估计. 香港股票市场看上去比日本股票市场有更大波动性，但是日本市场显示了波动率递增的趋势. 香港市场无条件的新息方差大约是 1.76，日本市场是 1.31.

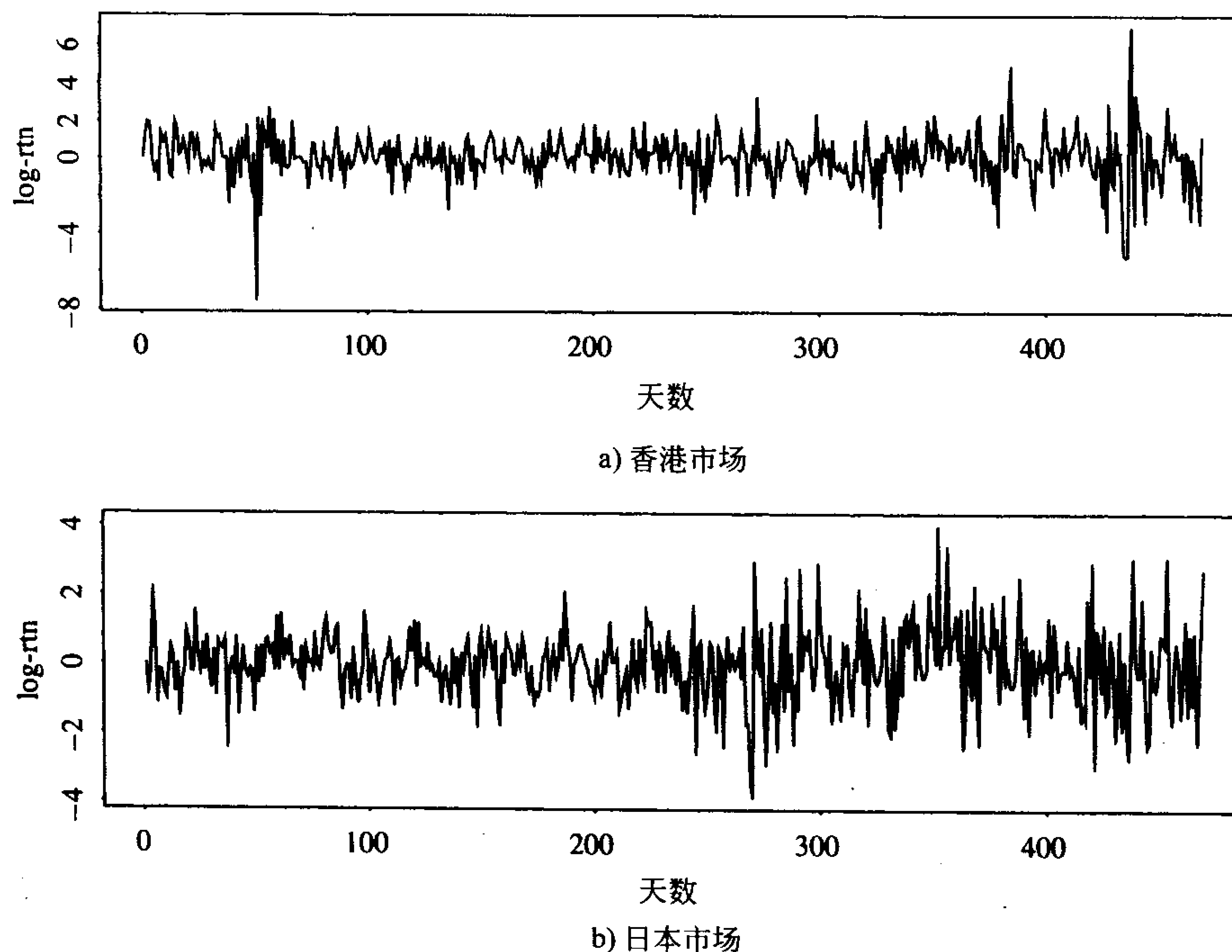


图 9-1　从 1996 年 1 月 1 日到 1997 年 10 月 16 日香港和日本股票市场指数的日对数收益率的时间图，收益率是以百分比给出的

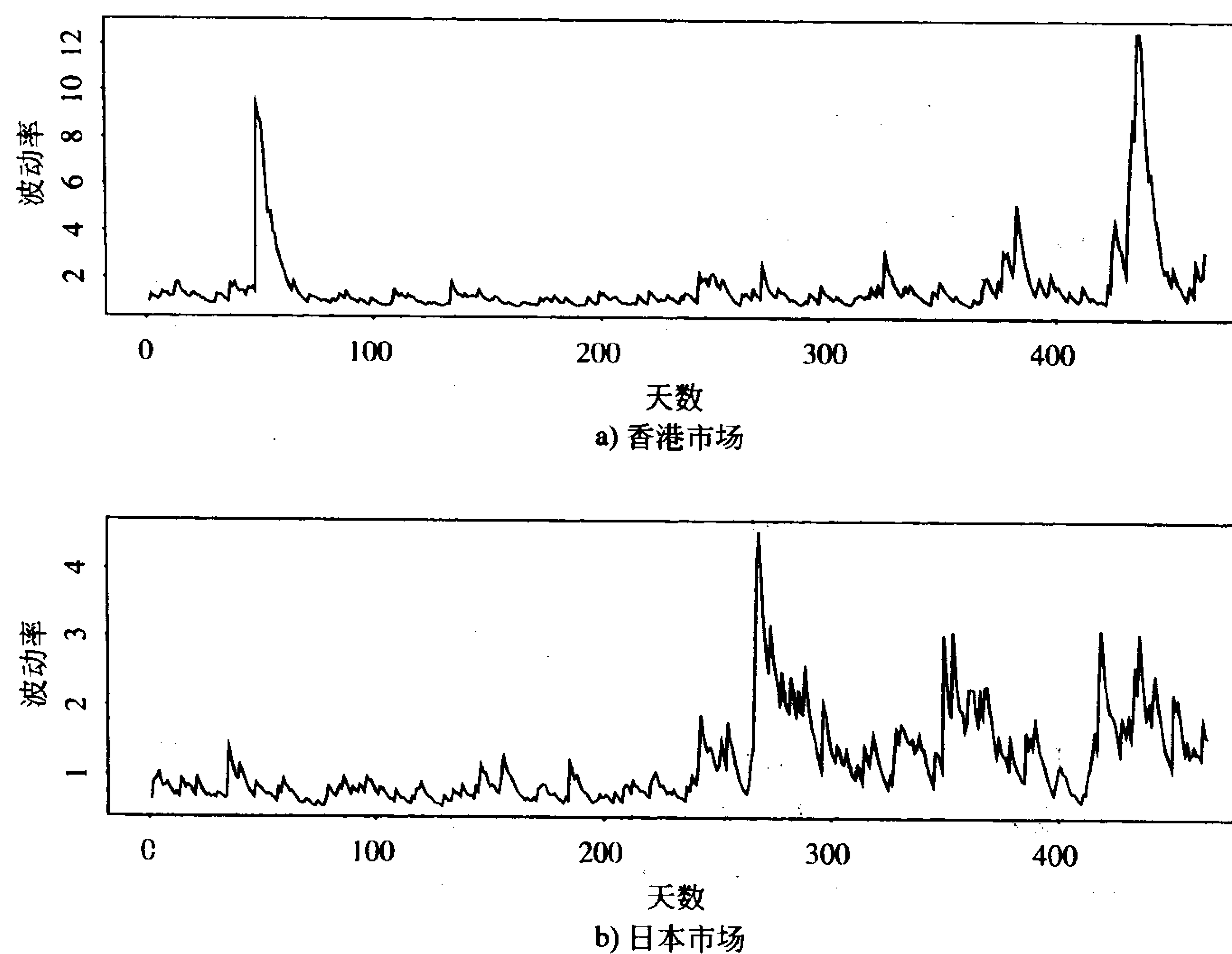


图 9-2　从 1996 年 1 月 1 日到 1997 年 10 月 16 日香港和日本股票市场指数的日对数收益率的波动率的估计值，收益率是以百分比给出的. 采用一元模型

365 ~ 366

进而考虑二元 GARCH 模型，我们得到了对数据拟合较好的两个模型. 第一个二元模型的均值方程为

$$r_{1t}=-0.118r_{1,t-6}+a_{1t},\quad r_{2t}=a_{2t},$$

其中 AR(6)系数的标准误差为 0.044. 第一个模型的波动率方程为

$$\begin{bmatrix}\sigma_{11,t}\\ \sigma_{22,t}\end{bmatrix}=\begin{bmatrix}0.275\\ (0.079)\\ 0.051\\ (0.014)\end{bmatrix}+\begin{bmatrix}0.112 & \cdot\\ (0.032) & \\ \cdot & 0.091\\ & (0.026)\end{bmatrix}\begin{bmatrix}a_{1,t-1}^2\\ a_{2,t-1}^2\end{bmatrix}$$

$$+\begin{bmatrix}0.711 & \cdot\\ (0.068) & \\ \cdot & 0.869\\ & (0.028)\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1}\\ \sigma_{22,t-1}\end{bmatrix}\tag{9.20}$$

其中小括号内的数为标准误差. a_{1t}与a_{2t}之间的相关系数的估计为 0.226，标准误差为 0.047.

令$\tilde{\boldsymbol{a}}_t=(\tilde{a}_{1t},\ \tilde{a}_{2t})'$是标准化的残差，其中$\tilde{a}_{it}=a_{it}/\sqrt{\sigma_{ii,t}}$. $\tilde{a}_t$ 的 L-B 统计量为 $Q(4)=22.29$ (0.10)，$Q(8)=34.83$(0.29)，括号内的数表示 p 值. 因为均值方程中利用了 AR(6)系数，所以这里的 p 值分别基于自由度为 15 和 31 的 χ^2 分布，$\tilde{a}_t^2$过程的 L-B 统计量为$Q(4)=9.54$ (0.85)，$Q(8)=18.58$(0.96). 因此，在模型(9.20)的二元标准化残差中没有序列相关性或条件异方差性. 香港市场和日本市场的两个残差的无条件新息方差分别是 1.55 和 1.28.

方程(9.20)的模型给出了两个分离的波动率方程，表明了两个市场的波动率不是动态相关的，但是它们是同步相关的. 我们将这种模型称为一个二元的对角常相关模型.

第二个二元 GARCH 模型的均值方程为

$$r_{1t}=-0.143r_{1,t-6}+a_{1t},\quad r_{2t}=a_{2t},$$

其中 AR(6)系数的标准误差为 0.042. 第二个模型的波动率方程为

367

$$\begin{bmatrix}\sigma_{11,t}\\ \sigma_{22,t}\end{bmatrix}=\begin{bmatrix}0.378\\ (0.103)\\ \cdot\\ \ \end{bmatrix}+\begin{bmatrix}0.108 & \cdot\\ (0.030) & \\ \cdot & 0.172\\ & (0.035)\end{bmatrix}\begin{bmatrix}a_{1,t-1}^2\\ a_{2,t-1}^2\end{bmatrix}$$

$$+\begin{bmatrix}\cdot & 0.865\\ & (0.109)\\ 0.321 & 0.869\\ (0.135) & (0.028)\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1}\\ \sigma_{22,t-1}\end{bmatrix},\tag{9.21}$$

括号内的数是标准误差. a_{1t}与 a_{2t}之间的相关系数的估计为 0.236，标准误差为 0.045. 如前一样定义标准化的残差，对前面模型的标准化误差，我们得到 $Q(4)=24.22(0.06)$，$Q(8)=35.52(0.26)$，由标准化残差的平方得到 $Q(4)=17.45(0.29)$，$Q(8)=24.55(0.79)$. 这些 L-B 统计量在 5%水平下都是不显著的，从而方程(9.21)中的模型也是合理的. 前面模型中香港和日本市场的新息的无条件方差分别为 1.71 和 1.32.

与模型(9.20)比较，第 2 个二元 GARCH(1，1)模型显示了两个市场之间的一种反馈关系，从而比较这两个波动率模型是很有趣的. 第一，模型(9.21)的新息的无条件方差接近于方程(9.18)和方程(9.19)的一元模型的结果. 第二，图 9-3 给出了模型(9.20)拟合的波动率过程，而图 9-4 显示了模型(9.21)拟合的过程. 因为模型(9.20)隐含了两个市场之间没有动态波动率依赖，图 9-3 类似于图 9-2. 相反，图 9-4 却给出了两个市场之间相互影响的证据. 第三，模型(9.20)的对数似然函数最大值为-535.13，其中 $t=8$，…，469，而模型(9.21)的值是-540.32；见方程(9.6)中的对数 pdf. 因此，如果用似然原理，则更喜欢用模型(9.20). 最后，因为两个二元波动率模型的实际含义有较大不同，为了将它们分离需要做进一步的研究. 这样的研究可能要使用一个更长的样本时期或包含更多的变量(例如使用一些 U.S. 市场的收益率).

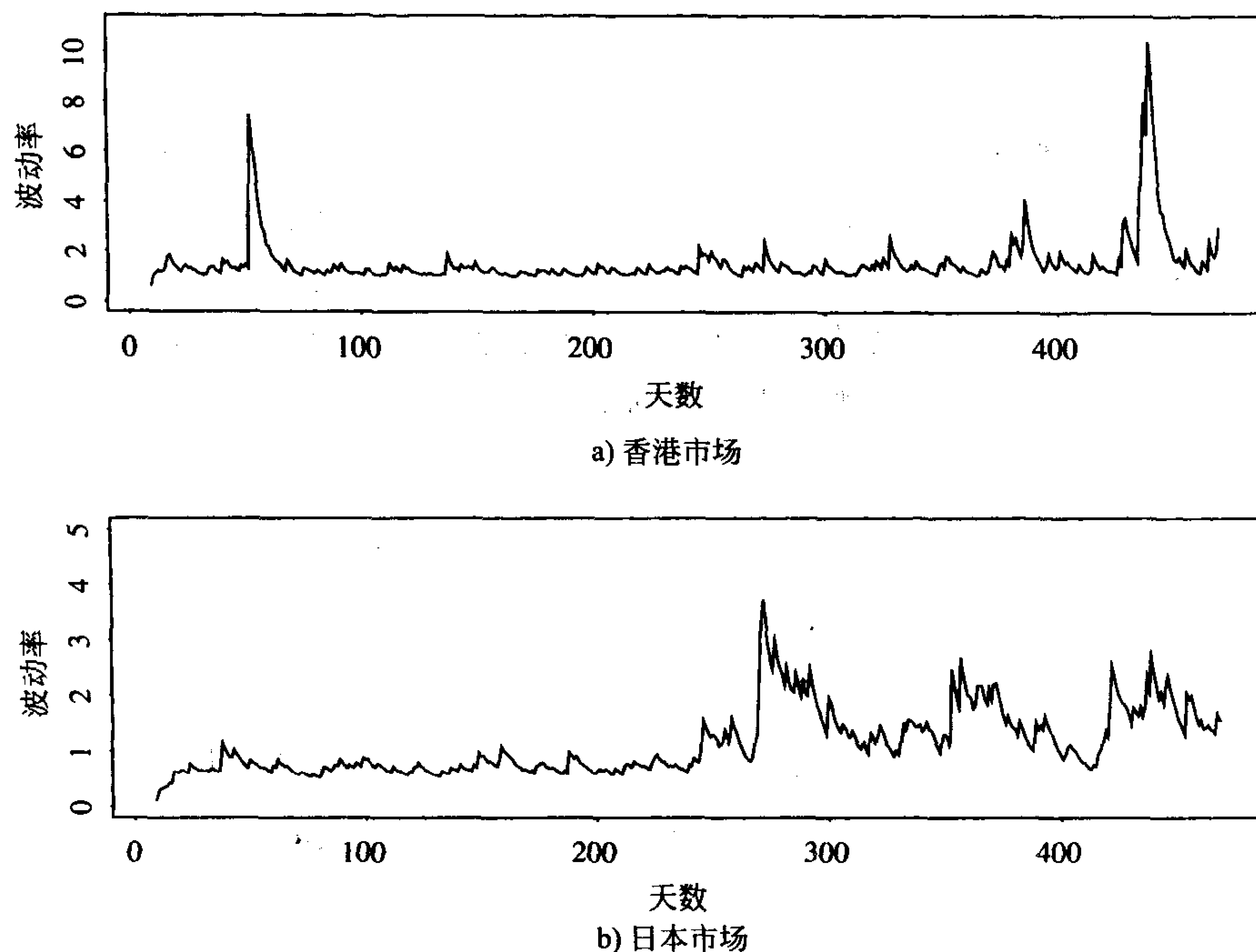


图 9-3 从 1996 年 1 月 1 日到 1997 年 10 月 16 日香港和日本股票市场指数的日对数收益率的波动率的估计值. 所用模型为方程(9.20)

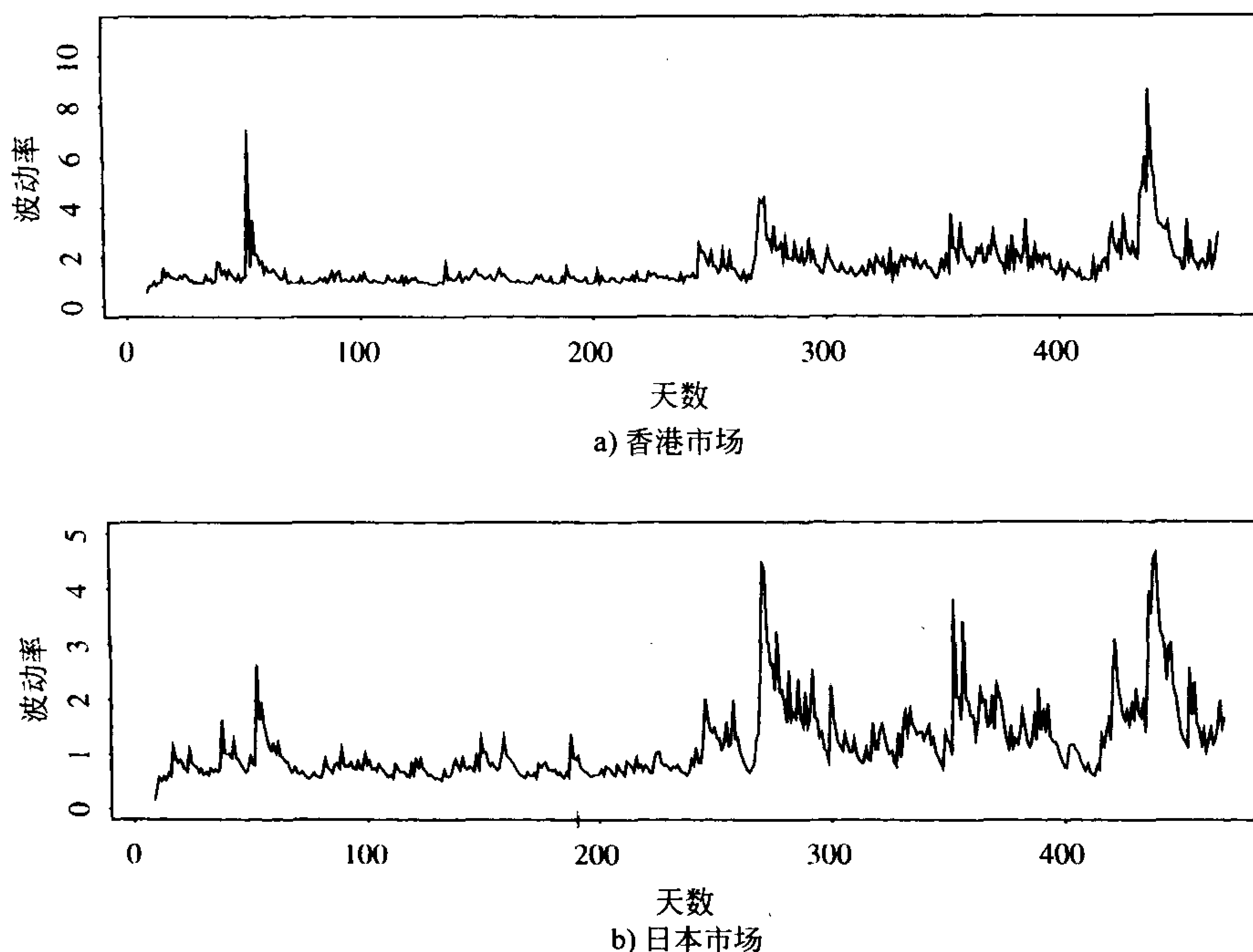


图 9-4　从 1996 年 1 月 1 日到 1997 年 10 月 16 日香港和日本股票市场指数的日对数收益率的波动率的估计值，收益率是以百分比给出的．所用模型为方程(9.21)

例 9.2　作为第二个解释，考虑第 8 章使用的 IBM 股票和标准普尔 500 指数从 1926 年 1 月至 1999 年 12 月以百分比表示的月对数收益率．令 r_{1t}和 r_{2t}分别表示 IBM 股票和标准普尔 500 指数的月对数收益率．如果利用一个常相关的 GARCH(1，1)模型，则我们得到的均值方程为

$$r_{1t}=1.351+0.072r_{1,t-1}+0.055r_{1,t-2}-0.119r_{2,t-2}+a_{1t},$$
$$r_{2t}=0.703+a_{2t},$$

第一个方程中参数的标准误差分别为 0.225，0.029，0.034 和 0.044，第二个方程中参数的标准误差为 0.155．波动率方程是

$$\begin{bmatrix}\sigma_{11,t}\\ \sigma_{22,t}\end{bmatrix}=\begin{bmatrix}2.98\\ (0.59)\\ 2.09\\ (0.47)\end{bmatrix}+\begin{bmatrix}0.079 & \cdot\\ (0.013) & \\ 0.042 & 0.045\\ (0.009) & (0.010)\end{bmatrix}\begin{bmatrix}a^2_{1,t-1}\\ a^2_{2,t-1}\end{bmatrix}+\begin{bmatrix}0.873 & -0.031\\ (0.020) & (0.009)\\ -0.066 & 0.913\\ (0.015) & (0.014)\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1}\\ \sigma_{22,t-1}\end{bmatrix}\qquad(9.22)$$

括号内的数为标准误差．常相关系数是 0.614，标准误差为 0.020．利用标准化残差，我们得到 L-B 统计量 $Q(4)=16.77(0.21)$，$Q(8)=32.40(0.30)$，括号内的 p 值分别是从自由度为 13 和 29 的 χ^2 分布得到的．这里的自由度已经调整了，因为均值方程包含三个延迟预测．对标准化残差的平方，我们有 $Q(4)=18.00(0.16)$，$Q(8)=39.09(0.10)$．因此，在 5%水平下，标准化的残差$\tilde{a}_t$ 没有序列相关性或条件异方差性．这个二元 GARCH(1，1)模型反映了这两个月对数收益率的波动率之间的反馈关系．

9.2.2 时变相关模型

常相关系数波动率模型的一个主要缺点是：在实际应用中，相关系数倾向于是时变的．考虑例 9.2 中使用的 IBM 股票和标准普尔 500 指数的月对数收益率．我们很难有理由认为加权平均的标准普尔 500 指数收益率和在过去 70 多年的 IBM 收益率保持一个常相关系数．图 9-5 利用 120 个(即 10 年)观测值的一个滑动窗口，给出了两个月对数收益率序列之间的样本相关系数．这个相关性是随时间变化的，且近年来看上去正在递减．相关性的递减趋势并不奇怪，因为 IBM 市场资本化在较大的美国．工业公司中的排名近年来已经改变．Tse(2000)提出了一个拉格朗日乘子统计量来检验多元 GARCH 模型中的常相关系数的假定．

368～370

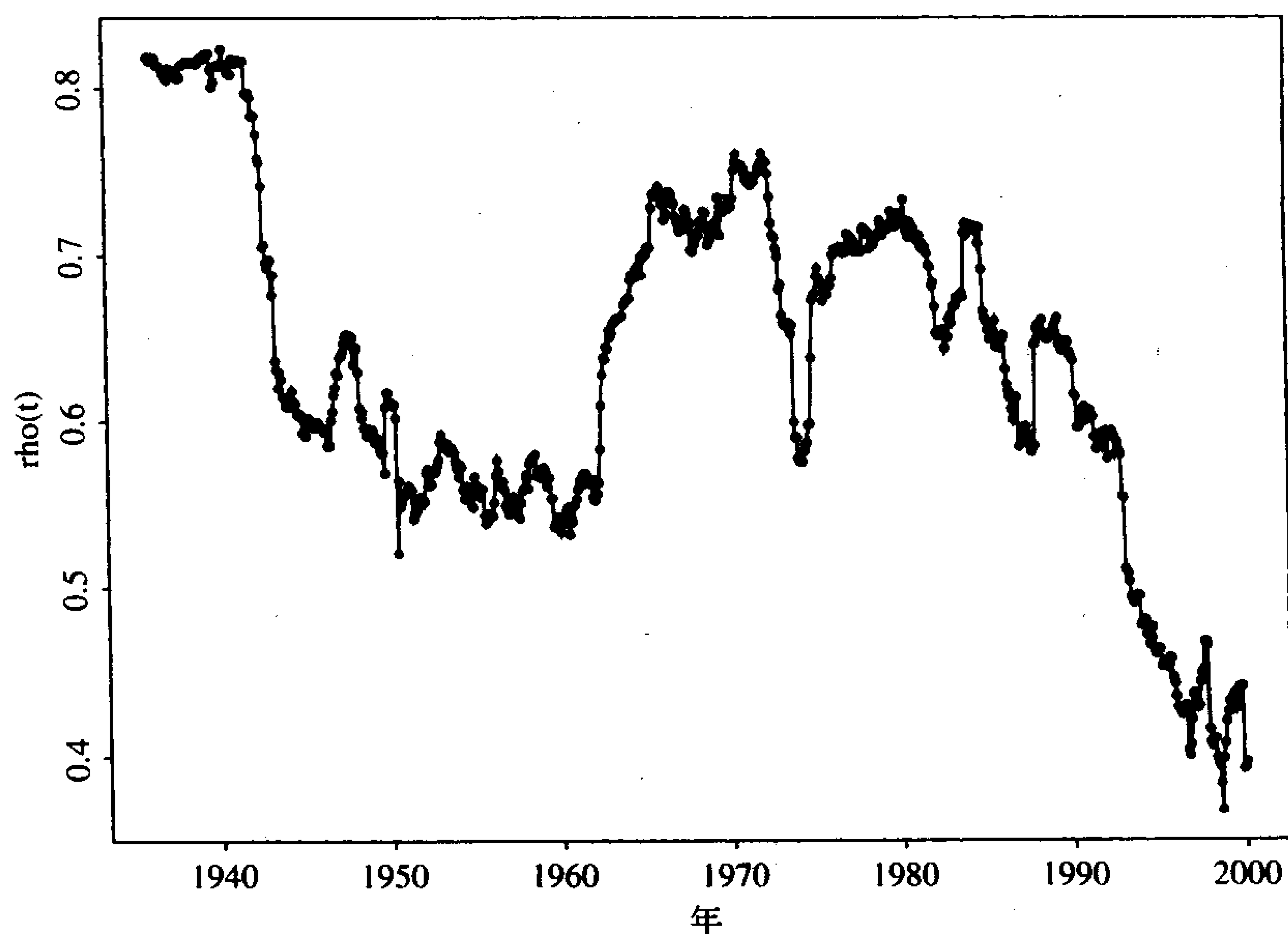


图 9-5 IBM 股票与标准普尔 500 指数的月对数收益率的样本相关系数．这个相关系数是由 120 个观察值的移动窗口计算出来的．样本时间区间是 1926 年 1 月至 1999 年 12 月

在 GARCH 框架内放松常相关限制的一个简单方法是对条件相关系数指定一个精确方程. 这可以利用 9.1 节讨论的Σ_t的两个重新参数化，通过两种方法来进行. 首先，我们直接利用相关系数. 因为 IBM 股票和标准普尔 500 指数收益率间的相关系数为正，而且一定在区间[0，1]中，所以我们用方程

$$\rho_{21,t}=\frac{\exp(q_t)}{1+\exp(q_t)},\tag{9.23}$$

其中 $q_t=\bar{\omega}_0+\bar{\omega}_1\rho_{21,t-1}+\bar{\omega}_2\frac{a_{1,t-1}a_{2,t-1}}{\sqrt{\sigma_{11,t-1}\sigma_{22,t-1}}}$，$\sigma_{ii,t-1}$是抖动 $a_{i,t-1}$的条件方差. 我们将这个方程称为相关系数的一个 GARCH(1，1)模型，因为它利用了两个抖动的 1 步延迟交叉相关系数与 1 步延迟交叉乘积. 如果 $\bar{\omega}_1=\bar{\omega}_2=0$，则模型(9.23)退化为常相关的情形.

371

总之，时变相关的二元 GARCH(1，1)模型包含了两组方程. 第一组方程包含了条件方差的一个二元 GARCH(1，1)模型，第二组方程是方程(9.23)中相关系数的一个 GARCH(1，1)模型. 实际中，如果相关系数为负，则可以在方程(9.23)中添加一个负号. 一般地，当相关性的符号未知时，我们可以利用相关系数的费希尔变换(Fisher Transformation).

$$q_t=\ln\left(\frac{1+\rho_{21,t}}{1-\rho_{21,t}}\right),\quad 或\ \rho_{21,t}=\frac{\exp(q_t)-1}{\exp(q_t)+1},$$

并对 q_t 采用一个 GARCH 模型来对两个收益率之间的时变相关系数建模.

例 9.2(续)　对 IBM 股票和标准普尔 500 指数的月对数收益率建立方程(9.22)中的 GARCH(1，1)模型，将方程(9.23)增加到其中，并进行联合估计，则对两个序列我们得到下面的模型：

$$r_{1t}=1.318+0.076r_{1,t-1}-0.068r_{2,t-2}+a_{1t},$$
$$r_{2t}=0.673+a_{2t}.$$

第一个方程中三个参数的标准误差分别为 0.215，0.026 和 0.034，第二个方程中参数的标准误差为 0.151. 波动率方程为

$$\begin{bmatrix}\sigma_{11,t}\\ \sigma_{22,t}\end{bmatrix}=\begin{bmatrix}2.80\\(0.58)\\1.71\\(0.40)\end{bmatrix}+\begin{bmatrix}0.084 & \cdot\\(0.013) & \\0.037 & 0.054\\(0.009) & (0.010)\end{bmatrix}\begin{bmatrix}a^2_{1,t-1}\\a^2_{2,t-1}\end{bmatrix}+\begin{bmatrix}0.864 & -0.020\\(0.021) & (0.009)\\-0.058 & 0.914\\(0.014) & (0.013)\end{bmatrix}\begin{bmatrix}\sigma_{11,t-1}\\ \sigma_{22,t-1}\end{bmatrix}\tag{9.24}$$

如前，括号内的数为标准误差. 条件相关系数方程为

$$\rho_t = \frac{\exp(q_t)}{1+\exp(q_t)}, \quad q_t = -2.024 + 3.983\rho_{t-1} + 0.088\frac{a_{1,t-1}a_{2,t-1}}{\sqrt{\sigma_{11,t-1}\sigma_{22,t-1}}}, \tag{9.25}$$

估计的标准误差分别为 0.050、0.090 和 0.019. 前面相关系数方程的参数是高度显著的. 对标准化残差$\tilde{a}_t$应用 L-B 统计量，有 $Q(4)=20.57(0.11)$，$Q(8)=36.08(0.21)$. 对标准化残差的平方，有 $Q(4)=16.69(0.27)$，$Q(8)=36.71(0.19)$. 因此，模型的标准化残差没有显著的序列相关或条件异方差性.

372

将这个时变相关系数的 GARCH(1，1)模型与方程(9.22)中的常相关系数的 GARCH(1，1)模型比较是很有趣的. 首先，两个模型的均值方程和波动率方程都接近. 其次，图 9-6 给出了基于模型(9.25)对 IBM 股票和标准普尔 500 指数的月对数收益率所拟合的条件相关系数. 此图表明该相关系数是随时间波动的，而且近几年变得更小. 后一个特征与图 9-5 的结论一致. 第三，对相关系数拟合的平均值为 0.612，这与方程(9.22)中常相关模型的估计 0.614 基本相同. 第四，利用 r_{it}的样本方差作为条件方差的初始值，观测值是从 $t=4$ 到 $t=888$，则对常相关 GARCH(1，1)模型，其对数似然函数的最大值为 −3 691.21，对时变相关 GARCH(1，1)模型，其对数似然函数的最大值为 −3 679.64. 这表明，时变相关模型比常相关模型有一些显著的改善. 最后，考虑两个模型在预测原点 $h=888$ 的 1 步向前波动率预测. 对方程(9.22)的常相关模型，我们有 $a_{1,888}=3.075$，$a_{2,888}=4.931$，$\sigma_{11,888}=77.91$，$\sigma_{22,888}=21.19$. 因此，对条件协方差矩阵的 1 步向前预测为

$$\hat{\Sigma}_{888}(1)=\begin{bmatrix}71.09 & 21.83\\ 21.83 & 17.79\end{bmatrix},$$

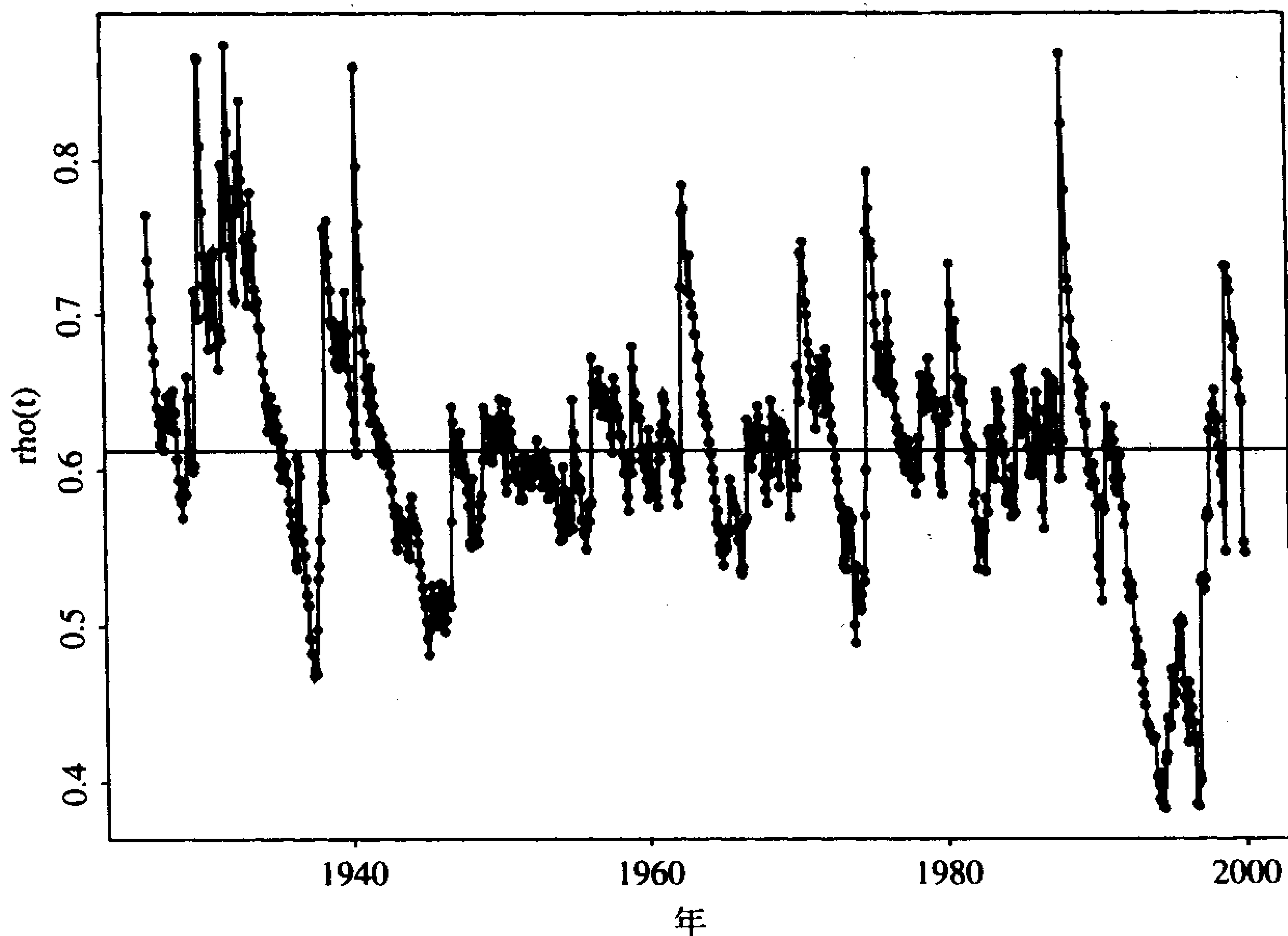


图 9-6　用例 9.2 中时变相关系数 GARCH(1，1)模型拟合的 IBM 股票与标准普尔 500 指数的月对数收益率的条件相关系数. 水平线是相关系数的平均值 0.612

373

其中协方差是通过常相关系数 0.614 得到的．对于方程(9.24)和方程(9.25)的时变相关模型，我们有 $a_{1,888}=3.287$，$a_{2,888}=4.950$，$\sigma_{11,888}=83.35$，$\sigma_{22,888}=28.56$，$\rho_{888}=0.546$．协方差矩阵的 1 步向前预测为

$$\hat{\Sigma}_{888}(1)=\begin{bmatrix}75.15 & 23.48\\ 23.48 & 24.70\end{bmatrix},$$

其中相关系数的预测为 0.545.

在第二个方法中，我们利用 Σ_t 的楚列斯基分解对时变相关性建模．对二元情形，参数向量为 $\boldsymbol{\Xi}_t=(g_{11,t}, g_{22,t}, q_{21,t})'$；见方程(9.13)．$\boldsymbol{a}_t$ 的一个简单的 GARCH(1，1)类型的模型为

$$\begin{aligned}
g_{11,t} &= \alpha_{10}+\alpha_{11}b_{1,t-1}^2+\beta_{11}g_{11,t-1},\\
q_{21,t} &= \gamma_0+\gamma_1 q_{21,t-1}+\gamma_2 a_{2,t-1},\\
g_{22,t} &= \alpha_{20}+\alpha_{21}b_{1,t-1}^2+\alpha_{22}b_{2,t-1}^2+\beta_{21}g_{11,t-1}+\beta_{22}g_{22,t-1},
\end{aligned}\tag{9.26}$$

其中，$b_{1t}=a_{1t}$，$b_{2t}=a_{2t}-q_{21,t}a_{1t}$．这样，假定 b_{1t} 为一个一元 GARCH(1，1)模型，b_{2t} 使用一个二元 GARCH(1，1)模型，$q_{21,t}$ 是自相关的，并利用 $a_{2,t-1}$ 作为一个额外的解释变量．与 MLE 相关的概率密度函数由方程(9.15)给出，取 $k=2$.

例 9.2(续)　我们再次使用 IBM 股票和标准普尔 500 指数的月对数收益率来说明方程(9.26)的波动率模型．利用与前面同样的指定，我们得到拟合的均值方程为

$$\begin{aligned}
r_{1t} &= 1.364+0.075r_{1,t-1}-0.058r_{2,t-2}+a_{1t}\\
r_{2t} &= 0.643+a_{2t}
\end{aligned}$$

第一个方程中参数的标准误差分别为 0.219，0.027 和 0.032，第二个方程中参数的标准误差为 0.154．这两个均值方程都接近于以前得到的结果．拟合的波动率模型为

$$\begin{aligned}
g_{11,t} &= 3.714+0.113b_{1,t-1}^2+0.804g_{11,t-1},\\
q_{21,t} &= 0.0029+0.9915q_{21,t-1}-0.0041a_{2,t-1},\\
g_{22,t} &= 1.023+0.021b_{1,t-1}^2+0.052b_{2,t-1}^2-0.040g_{11,t-1}+0.937g_{22,t-1},
\end{aligned}\tag{9.27}$$

[374]

其中 $b_{1t}=a_{1t}$，$b_{2t}=a_{2t}-q_{21,t}a_{1t}$．方程 $g_{11,t}$ 中参数的标准误差为 1.033，0.022 和 0.037，方程 $q_{21,t}$ 中参数的标准误差分别为 0.001，0.002 和 0.000 4，方程 $g_{22,t}$ 中参数的标准误差分别为 0.344，0.007，0.013 和 0.015．所有的估计都在 1%水平下统计显著.

条件协方差 Σ_t 可以利用方程(9.7)中的楚列斯基分解从模型(9.27)中得到．对二元情形，这个关系在方程(9.8)中已经详细给出．因此，我们得到时变相关系数为

$$\rho_t=\frac{\sigma_{21,t}}{\sqrt{\sigma_{11,t}\sigma_{22,t}}}=\frac{q_{21,t}\sqrt{g_{11,t}}}{\sqrt{g_{22,t}+q_{21,t}^2 g_{11,t}}}.\tag{9.28}$$

利用 $\sigma_{11,t}$ 和 $\sigma_{22,t}$ 的拟合值，我们可以计算标准化残差来进行模拟检验．模型(9.27)的标准化残差的 L-B 统计量为 $Q(4)=19.77(0.14)$，$Q(8)=34.22(0.27)$．对标准化残差的平方，我们有 $Q(4)=15.34(0.36)$，$Q(8)=31.87(0.37)$．这样，拟合的模型在描述条件均值和波动率方面

是合适的. 这个模型反映了该相关性有很强的动态依赖；这可从方程(9.27)的系数 0.991 5 看出.

图 9-7 给出了方程(9.28)拟合的时变相关系数，它给出了比图 9-6 更光滑的相关模式，并确认了相关系数的递减趋势. 特别地，所拟合的最近几年的相关系数要小于其他模型的拟合值. 对 IBM 股票和标准普尔 500 指数的月对数收益率的两个时变相关模型具有可比较的最大化的似然函数值，大约为−3 672，这表明拟合是相似的. 然而，基于楚列斯基分解的方法可能有一些优点. 首先，它并不要求为了保证Σ_t 的正定性而在估计中做任何参数限制. 如果也利用了 $g_{ii,t}$的对数变换，则对整个的波动率模型都不需要限制. 其次，在变换下，对数似然函数变得很简单. 第三，时变参数 $q_{ij,t}$和 $g_{ii,t}$有良好的解释. 然而，该变换使得推断有点复杂，因为拟合的模型可能依赖于 $\boldsymbol{a}_t$ 中元素的次序；记住，a_{1t}没有变换. 理论上，$\boldsymbol{a}_t$ 中元素的次序应该不会对波动率产生影响.

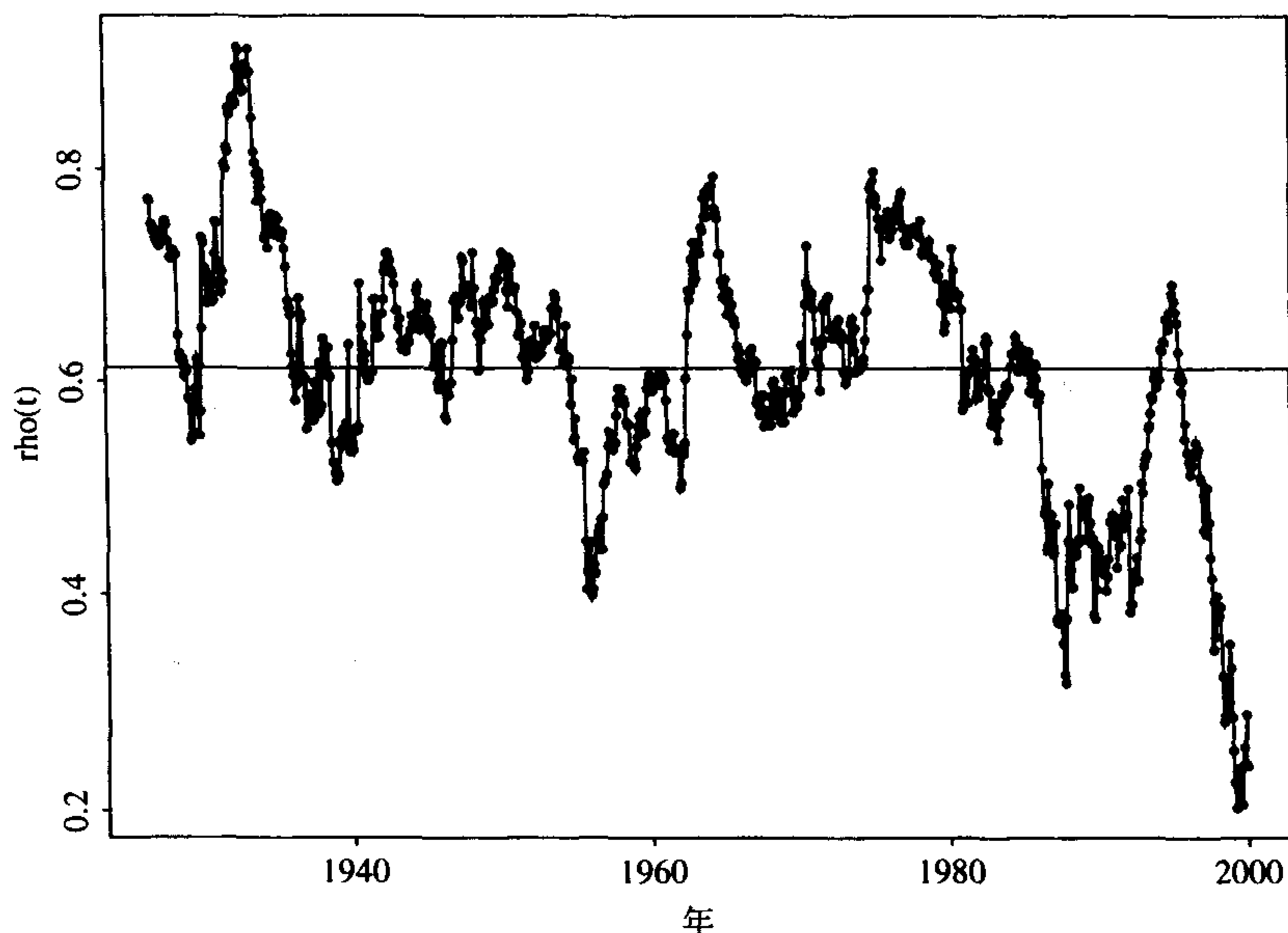


图 9-7 用带楚列斯基分解的时变相关系数 GARCH(1, 1)模型拟合的 IBM 股票与标准普尔 500 指数的月对数收益率的条件相关系数. 水平线是相关系数估计的平均值 0.612

最后，对于新的时变相关模型，条件协方差矩阵在预测原点 $t=888$ 处的 1 步向前预测为

$$\hat{\Sigma}_{888}(1)=\begin{bmatrix}73.45 & 7.34\\ 7.34 & 17.87\end{bmatrix}.$$

由此得到的相关系数的预测为 0.203，它远小于前两个模型所得的结果. 然而，条件方差的预测与前面类似.

注释： *在最近的一个原稿中，Tse 和 Tsui(1998)考虑了一个具有时变相关性的多元 GARCH 模型. 对 k 维收益率，作者假定条件相关矩阵$\boldsymbol{\rho}_t$ 服从模型*

$$\boldsymbol{\rho}_t = (1-\theta_1-\theta_2)\boldsymbol{\rho} + \theta_1 \boldsymbol{\rho}_{t-1} + \theta_2 \psi_{t-1},$$

其中 θ_1 和 θ_2 是尺度参数，$\boldsymbol{\rho}$ 是具有单位对角元素的 $k\times k$ 正定矩阵，ψ_{t-1} 是 $k\times k$ 样本相关矩阵，它是利用时间点 $t-m$，…，$t-1$ 上的抖动计算的，其中 m 是事先指定的. 要求对两个尺度参数 θ_1 和 θ_2 的估计作特殊的限制以保证相关矩阵的正定性. 这个方法看起来比本章中考虑的两个方法显得更加复杂.

9.3 更高维的波动率模型

本节中，利用楚列斯基分解的序贯性，我们提出一个构造更高维波动率模型的策略. 再次将向量收益序列写为 $\boldsymbol{r}_t=\boldsymbol{\mu}_t+\boldsymbol{a}_t$. 可以利用第 8 章的方法为 $\boldsymbol{r}_t$ 指定一个均值方程. 通常简单的向量 AR 模型就足够了. 这里我们集中讨论利用抖动过程 $\boldsymbol{a}_t$ 建立波动率模型. 375~376

根据 9.1 节对楚列斯基分解的讨论，从 a_{it} 到 b_{it} 的正交变换只涉及了 $b_{jt}(j<i)$. 另外，从 $g_{ii,t}$ 的模型只依赖于与 $b_{jt}(j<i)$ 相关的量这种意义上讲，9.2 节建立的时变波动率模型是形成一族的. 因此，我们考虑按照下面的一系列程序来建立多元波动率模型.

1)选择一个主要感兴趣的市场指数或股票收益率. 对选择的收益率序列建立一个一元波动率模型.

2)将第二个收益率序列扩充到系统中，将扩充后的新收益率序列的抖动过程进行正交变换，并对系统建立一个二元波动率模型. 则第 1 步中的一元模型的参数估计可以作为二元估计的初始值.

3)将第三个收益率序列扩充到系统中，对新增加的抖动过程进行正交变换，并建立一个三维波动率模型. 二元模型的参数估计又作为三维估计的初始值.

4)继续此扩充直到对所有感兴趣的收益率序列建立了联合波动率模型.

最后，在每一步中进行模型检验以保证所拟合模型的合理性. 经验表明，这一系列程序可以从实质上简化建立高维波动率模型的复杂性. 特别是可以大大减少估计的计算时间.

例 9.3 我们通过对标准普尔 500 指数，Cisco 系统和 Intel 公司的股票的日对数收益率建立一个波动率模型来说明所提出的一系列程序. 数据的时间区间是从 1991 年 1 月 2 日至 1999 年 12 月 31 日，共 227 5 个观测值. 对数收益率是以百分比表示的，在图 9-8 中给出. 收益率序列的分量排序为 $\boldsymbol{r}_t=(\text{SP5}_t, \text{CSCO}_t, \text{INTC}_t)'$. 数据的样本均值、标准差和相关矩阵为

$$\hat{\boldsymbol{\mu}}=\begin{bmatrix}0.066\\0.257\\0.156\end{bmatrix},\quad \begin{bmatrix}\hat{\sigma}_1\\\hat{\sigma}_2\\\hat{\sigma}_3\end{bmatrix}=\begin{bmatrix}0.875\\2.853\\2.464\end{bmatrix},\quad \hat{\boldsymbol{\rho}}=\begin{bmatrix}1.00&0.52&0.50\\0.52&1.00&0.47\\0.50&0.47&1.00\end{bmatrix}.$$

利用 L-B 统计量来识别收益率序列之间的序列相关性，得到 $Q(1)=26.20$，$Q(4)=79.73$，$Q(8)=123.68$. 这些检验统计量都是高度显著的，分别与自由度为 9、36 和 72 的 χ^2 377

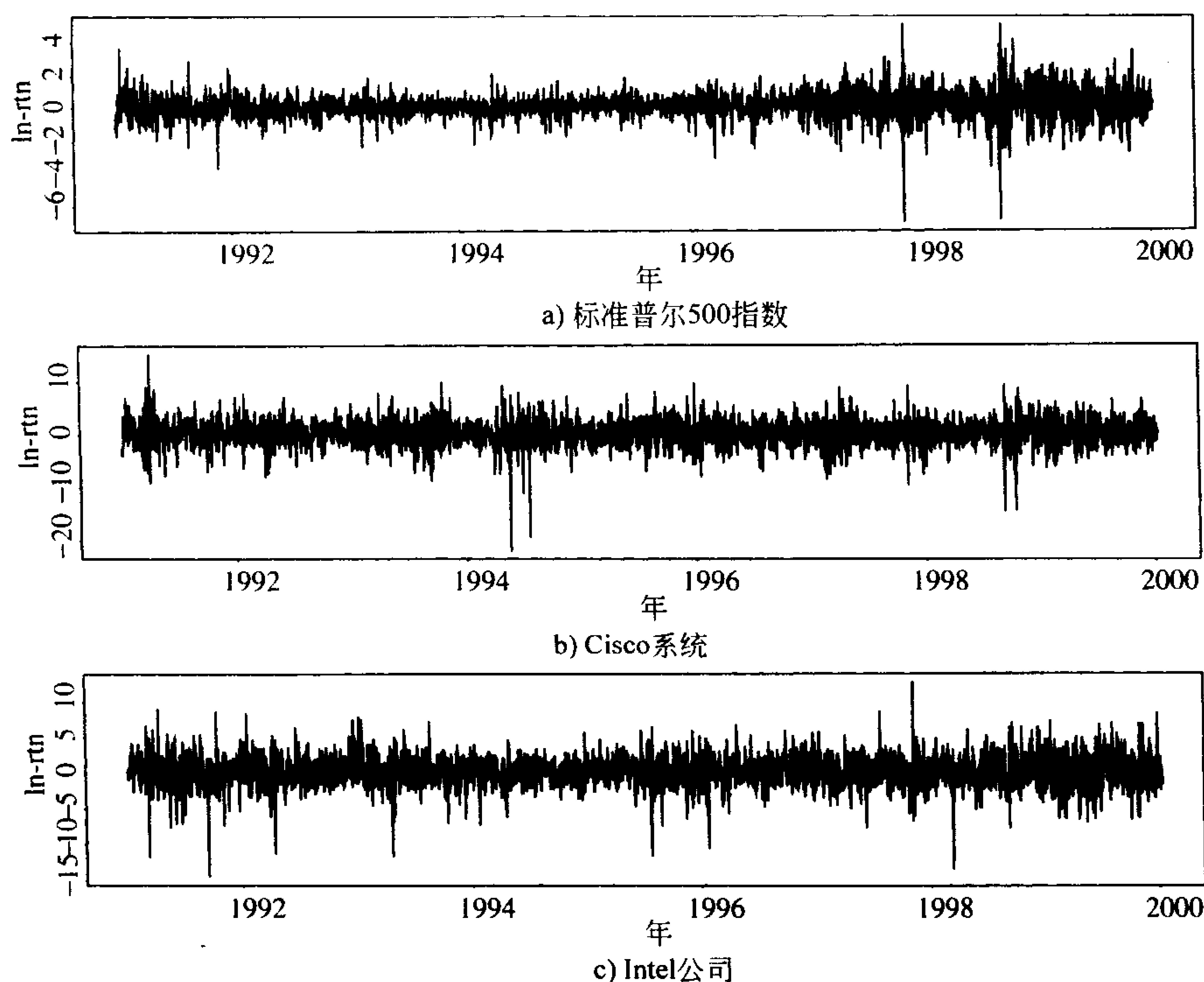


图 9-8 从 1991 年 1 月 2 日到 1999 年 12 月 31 日标准普尔 500 指数，Cisco 系统和 Intel 公司股票的日对数收益率（百分比形式）的时间图

分布比较，其 p 值都接近于 0. 数据中确实具有序列相关性. 表 9-1 给出了以第 8 章中的简单记号表示的样本交叉相关矩阵的前 5 个延迟. 检查此表可见(a)标准普尔 500 指数的日对数收益率并不依赖于 Cisco 或 Intel 过去的收益率. (b)Cisco 股票的对数收益率具有某些序列相关性，并且依赖于标准普尔 500 指数的过去收益率（见延迟 2 和 5）. (c)Intel 股票的对数收益率依赖于标准普尔 500 指数的过去收益率（见延迟 1 和 5）. 这些观测类似于第 8 章中分析的 IBM 股票和标准普尔 500 指数收益率之间的结果. 它们表明具有较大资本的单个公司的收益率倾向于受到市场过去行为的影响. 然而，市场收益率并不会受到单个公司过去收益率的显著影响.

进而考虑波动率模型，按照所建议的步骤，我们以标准普尔 500 指数的对数收益率开始，得到模型

$$r_{1t} = 0.078 + 0.042 r_{1,t-1} - 0.062 r_{1,t-3} - 0.048 r_{1,t-4} - 0.052 r_{1,t-5} + a_{1t},$$
$$\sigma_{11,t} = 0.013 + 0.092 a_{1,t-1}^2 + 0.894 \sigma_{11,t-1}, \tag{9.29}$$

均值方程中参数的标准误差分别为 0.016，0.023，0.020，0.022 和 0.020，波动率方程中参数的标准误差分别为 0.002，0.006 和 0.007. 一元模型的标准化残差及标准化残差平方序列的 L-B 统计量没有识别出数据中任何余留的序列相关性或条件异方差性. 事实上，对标准化残

差，我们有 $Q(10)=7.38(0.69)$，对其平方序列，$Q(10)=3.14(0.98)$.

将 Cisco 股票的日对数收益率扩充到系统中，我们建立一个二元模型，具有的均值方程为

$$r_{1t}=0.065-0.046r_{1,t-3}+a_{1t},$$

$$r_{2t}=0.325+0.195r_{1,t-2}-0.091r_{2,t-2}+a_{2t}, \tag{9.30}$$

其中所有的估计在 1%水平下都是统计上显著的. 利用楚列斯基分解的记号，我们得到波动率方程为

$$g_{11,t}=0.006+0.051b_{1,t-1}^2+0.943g_{11,t-1},$$

$$q_{21,t}=0.331+0.790q_{21,t-1}-0.041a_{2,t-1},$$

$$g_{22,t}=0.177+0.082b_{2,t-1}^2+0.890g_{22,t-1}, \tag{9.31}$$

其中 $b_{1t}=a_{1t}$，$b_{2t}=a_{2t}-q_{21,t}b_{1t}$，$g_{11,t}$方程中参数的标准误差分别为 0.001，0.005 和 0.006，$q_{21,t}$方程中参数的标准误差分别为 0.156，0.099 和 0.011，$g_{22,t}$方程中参数的标准误差分别为 0.029，0.008 和 0.011. 标准化残差的二元 L-B 统计量没有识别出任何余留的序列相关性与条件异方差性，从而二元模型是合适的. 与方程(9.29)比较，我们看出 r_{1t}的边际模型和一元模型之间的差别很小.

表 9-1

Lag					
1	2	3	4	5	6
. . .	. . .	− . .	. . .	− . .	. . .
. . .	. − .	. . .	. . .	− . .	. − .
− . .	. . .	. . .	. . .	− . .	. . .

注：从 1991 年 1 月 2 日到 1999 年 12 月 31 日标准普尔 500 指数，Cisco 系统和 Intel 公司的股票的日对数收益率的样本交叉相关矩阵.

下一步也是最后一步，是将 Intel 股票的日对数收益率扩充到系统中. 均值方程变为

$$r_{1t}=0.065-0.043r_{1,t-3}+a_{1t},$$
$$r_{2t}=0.326+0.201r_{1,t-2}-0.089r_{2,t-2}+a_{2t}, \tag{9.32}$$
$$r_{3t}=0.192-0.264r_{1,t-1}+0.059r_{3,t-1}+a_{3t}.$$

378～379

第一个方程中参数的标准误差分别为 0.016 和 0.017，第二个方程中参数的标准误差分别为 0.052，0.059 和 0.021，第三个方程中参数的标准误差分别为 0.050，0.057 和 0.022. 所有的估计都在大约 1%水平下是统计上显著的. 如所料想，r_{1t}和 r_{2t}的均值方程与二元情形中的基本相同.

三维时变波动率模型变得更加复杂，但仍然可以处理为

$$g_{11,t}=0.006+0.050b_{1,t-1}^2+0.943g_{11,t-1},$$

$$
\begin{aligned}
q_{21,t} &= 0.277 + 0.824 q_{21,t-1} - 0.035 a_{2,t-1}, \\
g_{22,t} &= 0.178 + 0.082 b_{2,t-1}^2 + 0.889 g_{22,t-1}, \\
q_{31,t} &= 0.039 + 0.973 q_{31,t-1} + 0.010 a_{3,t-1}, \\
q_{32,t} &= 0.006 + 0.981 q_{32,t-1} + 0.004 a_{2,t-1}, \\
g_{33,t} &= 1.188 + 0.053 b_{3,t-1}^2 + 0.687 g_{33,t-1} - 0.019 g_{22,t-1},
\end{aligned} \tag{9.33}
$$

其中 $b_{1t}=a_{1t}$，$b_{2t}=a_{2t}-q_{21,t}b_{1t}$，$b_{3t}=a_{3t}-q_{31,t}b_{1t}-q_{32,t}b_{2t}$. 表 9-2 给出了参数的标准误差. 除了 $q_{32,t}$ 方程中的常数项，所有的估计都在 5%水平下显著. 令 $\tilde{\boldsymbol{a}}_t=(a_{1t}/\hat{\sigma}_{1t},\ a_{2t}/\hat{\sigma}_{2t},\ a_{3t}/\hat{\sigma}_{3t})'$ 为标准化的残差序列，其中 $\hat{\sigma}_{it}=\sqrt{\hat{\sigma}_{ii,t}}$ 是对第 i 个收益率拟合的条件标准差. $\tilde{\boldsymbol{a}}_t$ 的 L-B 统计量给出 $Q(4)=34.48(0.31)$，$Q(8)=60.42(0.70)$，其中 χ^2 分布的自由度分别为 31 和 67，该自由度是将均值方程中使用的参数数量调整之后的. 对于标准残差序列的平方 $\tilde{\boldsymbol{a}}_t^2$，我们有 $Q(4)=28.71(0.58)$，$Q(8)=52.00(0.91)$，因此，所拟合的模型作为给条件均值和波动率建立的模型是合适的.

表 9-2

方程	标准误差				方程	标准误差		
$g_{11,t}$	0.001	0.005	0.006		$q_{21,t}$	0.135	0.086	0.010
$g_{22,t}$	0.029	0.009	0.011		$q_{31,t}$	0.017	0.012	0.004
$g_{33,t}$	0.407	0.015	0.100	0.008	$q_{32,t}$	0.004	0.013	0.001

注：给从 1991 年 1 月 2 日到 1999 年 12 月 31 日标准普尔 500 指数，Cisco 系统和 Intel 公司股票的日对数收益率(百分比形式)的三维波动率模型的参数估计值的标准差. 参数出现的顺序和式(9.33)的顺序一样.

方程(9.33)中的三维波动率模型有一些有趣的特征. 首先，它本质上是一个时变相关的 GARCH(1，1)模型，因为方程中只用到了 1 步延迟变量. 其次，标准普尔 500 指数日对数收益率的波动率不依赖于 Cisco 或 Intel 股票收益率过去的波动率. 第三，通过取楚列斯基分解的逆变换，Cisco 和 Intel 股票日对数收益率的波动率依赖于市场收益率过去的波动率；见 9.1 节给出的 Σ_t，$\boldsymbol{L}_t$ 和 $\boldsymbol{G}_t$ 的元素之间的关系. 第四，描述相关的量 $q_{ij,t}$ 有很高的持续性，具有很大的 AR(1)系数.

380

图 9-9 给出了数据拟合模型的波动率过程(即 $\hat{\sigma}_{ii,t}$). 指数收益率的波动率远远小于两个单只股票收益率的波动率. 图形也表明了指数收益率的波动率近年来已经增加，但是 Cisco 系统的收益率却并非如此. 图 9-10 给出了三种收益率序列之间的时变相关系数. 比较图 9-9 与图 9-10可得到有意思的结论. 它们表明当收益率波动时，两个收益率序列之间的相关系数是递增的. 这与国际股票市场指数间关系的实证研究结果是一致的，两个市场之间的相关性在金融危机期间倾向于递增.

方程(9.33)中的波动率模型包含两组方程. 第一组方程描述了条件方差的时间演变(即

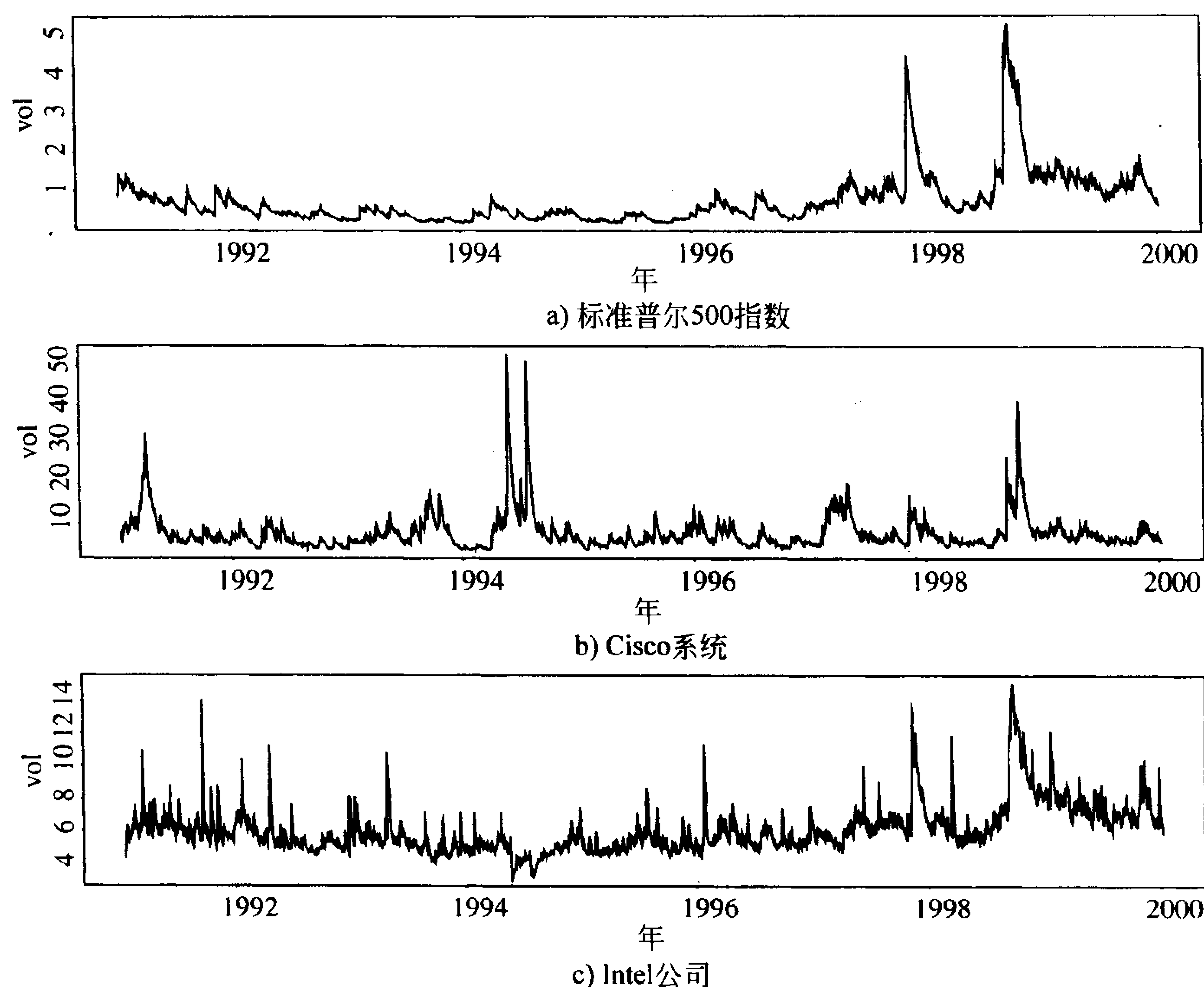


图 9-9 给从 1991 年 1 月 2 日到 1999 年 12 月 31 日标准普尔 500 指数，Cisco 系统和 Intel 公司股票的日对数收益率(百分比形式)拟合的波动率的时间图

$g_{ii,t}$)，第二组方程处理相关系数(即 $q_{ij,t}$，$i>j$)．对这个特殊的数据集合，一个 AR(1)模型对相关系数方程可能就足够了．类似地，一个简单的 AR 模型对条件方差也可能是足够的．定义 [381]

$$\boldsymbol{v}_t = (v_{11,t}, v_{22,t}, v_{33,t})',$$

这里 $v_{ii,t}=\ln(g_{ii,t})$，$\boldsymbol{q}_t=(q_{21,t},\ q_{31,t},\ q_{32,t})'$．前面的讨论启示我们可以利用简单的 1 步延迟模型 $\boldsymbol{v}_t=\boldsymbol{c}_1+\boldsymbol{\beta}_1\boldsymbol{v}_{t-1}$，$\boldsymbol{q}_t=\boldsymbol{c}_2+\boldsymbol{\beta}_2\boldsymbol{q}_{t-1}$，作为对资产收益率波动率建模的精确函数，其中 $\boldsymbol{c}_i$ 是常向量，$\boldsymbol{\beta}_i$ 是 3×3 实值矩阵．如果前一方程中再加上噪声项，则模型变为

$$\boldsymbol{v}_t = \boldsymbol{c}_1 + \boldsymbol{\beta}_1\boldsymbol{v}_{t-1} + \boldsymbol{e}_{1t}, \quad \boldsymbol{q}_t = \boldsymbol{c}_2 + \boldsymbol{\beta}_2\boldsymbol{q}_{t-1} + \boldsymbol{e}_{2t}.$$

其中 $\boldsymbol{e}_{it}$ 是均值为 0 的随机抖动，具有正定协方差矩阵，则我们有一个简单的多元随机波动率模型．在一个最近的手稿中，Chib，Nardarí 和 Shephard(1999)利用 MCMC 方法来研究高维的随机波动率模型．那里考虑的模型允许以一种带限制的方式时变的相关性．多元波动率模型的其他参考文献包括 Harvey，Ruiz 和 Shephard(1995)．在第 10 章中，我们讨论波动率建模的 MCMC 方法．

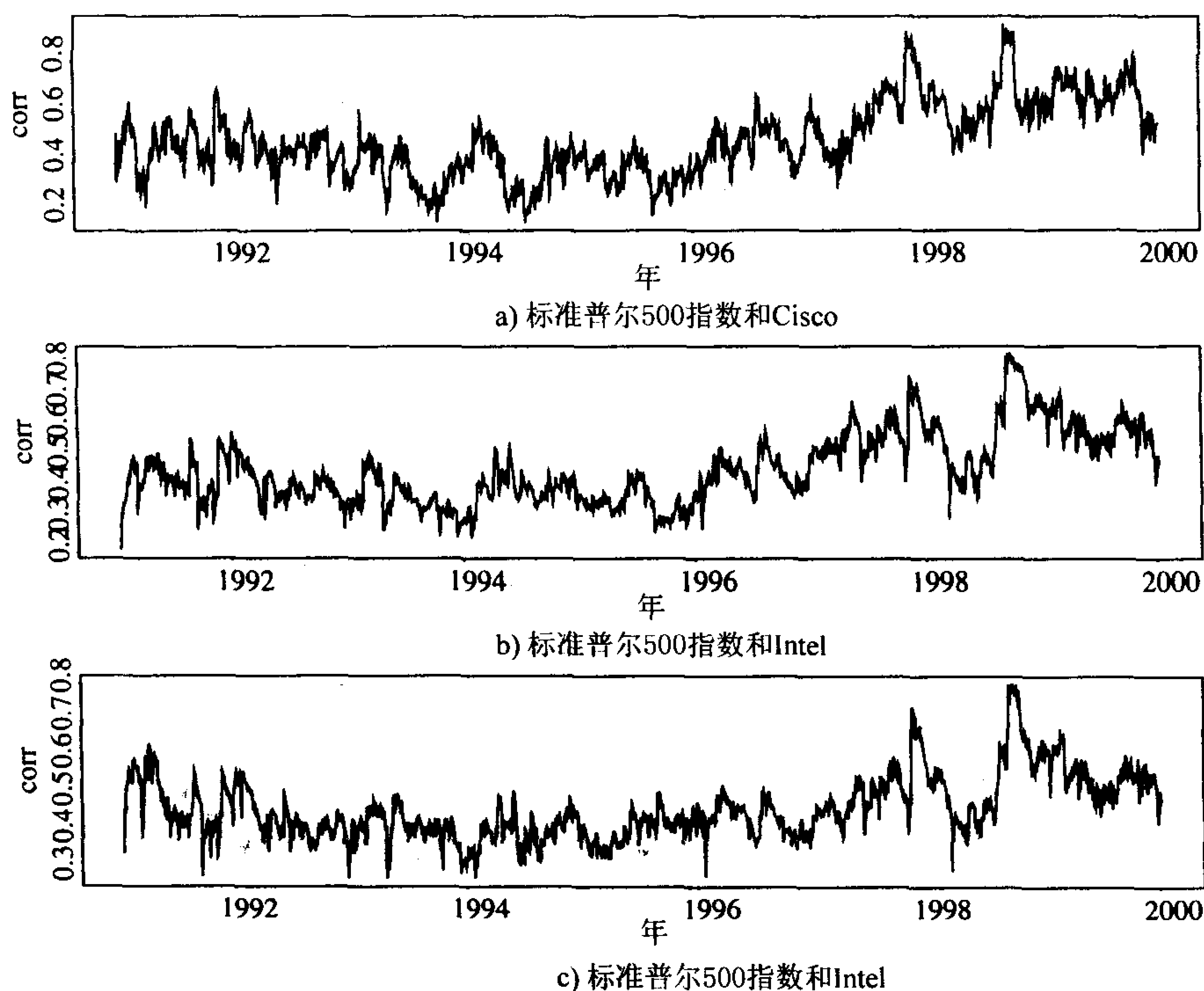


图 9-10 从 1991 年 1 月 2 日到 1999 年 12 月 31 日标准普尔 500 指数，Cisco 系统和 Intel 公司的股票的日对数收益率拟合的时变相关系数的时间图

382

9.4 因子波动率模型

简化多元波动率过程动态结构的另一种方法是利用因子模型. 实际中，“公共因子”可以通过固有性质或经验方法预先设定. 作为说明，我们利用第 8 章的因子分析来讨论因子波动率模型. 因为波动率模型关心的是 $\boldsymbol{a}_t$ 的条件方差矩阵随时间的演变，其中 $\boldsymbol{a}_t=\boldsymbol{r}_t-\boldsymbol{\mu}_t$，所以波动率中识别“公共因子”的一个简单方式是对 $\boldsymbol{a}_t$ 进行主成分分析(PCA)，见第 8 章中 PCA. 这样，建立一个因子波动率模型涉及一个三步的程序：

- 选择头几个的主成分，使得它们所能解释的 $\boldsymbol{a}_t$ 的变化占很高的百分比.
- 对已选择的主成分建立一个波动率模型.
- 对每个 a_{it} 序列的波动率与已选择的主成分的波动率联系起来.

这个程序的目的是降低维数，但保持多元波动率的较精确的近似.

例 9.4 再次考虑例 9.2 中的 IBM 股票和标准普尔 500 指数以百分比表示的月对数收益率. 利用例 8.4 中的二元 AR(3)模型，我们得到新息序列 $\boldsymbol{a}_t$. 根据 $\boldsymbol{a}_t$ 的协方差矩阵对 $\boldsymbol{a}_t$ 进行主成分分析，我们得到特征值 63.373 和 13.489. 第一个特征值解释了 $\boldsymbol{a}_t$ 的广义方差的

82.2%．因此，我们可以选择第一个主成分 $x_t=0.797a_{1t}+0.604a_{2t}$ 作为公共因子．另外一种选择，正如例 8.4 中的模型表明的，$\boldsymbol{r}_t$ 的序列相关性很弱，从而可以直接对 $\boldsymbol{r}_t$ 进行主成分分析．对这个特例，$\boldsymbol{r}_t$ 的样本协方差矩阵的两个特征值是 63.625 和 13.513，这与基于 $\boldsymbol{a}_t$ 的结果基本相同．第一个主成分近似解释了 $\boldsymbol{r}_t$ 的广义方差的 82.5%，对应的公共因子是 $x_t=0.796r_{1t}+0.605r_{2t}$．因此，对所考虑的两个月对数收益率序列，条件均值方程对 PCA 的影响是可以忽略的．

根据前面的讨论，并且为了简便，对这两个月对数收益率序列，我们使用 $x_t=0.796r_{1t}+0.605r_{2t}$ 作为一个公共因子．图 9-11a 给出了这个公共因子的时间图．如果使用一元高斯 GARCH 模型，则对 x_t 我们得到下面的模型：

383

$$x_t = 1.317 + 0.096x_{t-1} + a_t, \quad a_t = \sigma_t\epsilon_t,$$

$$\sigma_t^2 = 3.834 + 0.110a_{t-1}^2 + 0.825\sigma_{t-1}^2, \tag{9.34}$$

这个模型的所有参数估计在 1% 水平下都是高度显著的，而且标准化的残差及其平方序列的 L-B 统计量没有识别出模型的不合适性．图 9-11b 给出了 x_t 的拟合波动率(即方程(9.34)中的样本 σ_t^2 序列)．

利用模型(9.34)中的 σ_t^2 作为一个公共因子，对于原始的月对数收益率，我们得到下面的模型．均值方程为

$$r_{1t} = 1.140 + 0.079r_{1,t-1} + 0.067r_{1,t-2} - 0.122r_{2,t-2} + a_{1t},$$

$$r_{2t} = 0.537 + a_{2t},$$

第一个方程中参数的标准误差分别是 0.211，0.030，0.031 和 0.043，第二个方程中参数的标准误差是 0.165．条件方差方程为

$$\begin{bmatrix}\sigma_{11,t}\\ \sigma_{22,t}\end{bmatrix} = \begin{bmatrix}19.08\\ (3.70)\\ -5.62\\ (2.36)\end{bmatrix} + \begin{bmatrix}0.098 & \cdot\\ (0.044) & \\ \cdot & \cdot\end{bmatrix}\begin{bmatrix}a_{1,t-1}^2\\ a_{2,t-1}^2\end{bmatrix} + \begin{bmatrix}0.333\\ (0.076)\\ 0.596\\ (0.050)\end{bmatrix}\sigma_t^2 \tag{9.35}$$

如前一样，括号内的数是标准差，σ_t^2 是从模型(9.34)得到的．条件相关系数方程为

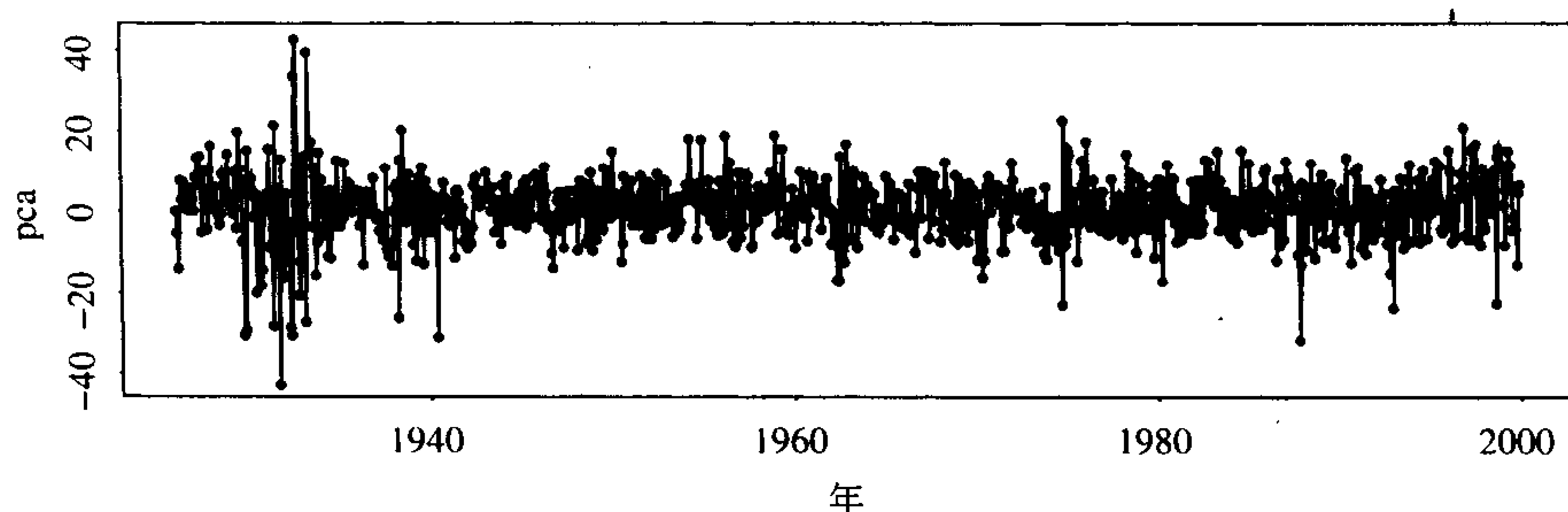
$$\rho_t = \frac{\exp(q_t)}{1+\exp(q_t)}, \quad q_t = -2.098 + 4.120\rho_{t-1} + 0.078\frac{a_{1,t-1}a_{2,t-1}}{\sqrt{\sigma_{11,t-1}\sigma_{22,t-1}}}. \tag{9.36}$$

三个参数的标准误差分别是 0.025，0.038 和 0.015．如前一样定义标准化的残差序列，得到 $Q(4)=15.37(0.29)$，$Q(8)=34.24(0.23)$，括号内的数表示 p 值．因此，标准化的残差没有序列相关性．对残差平方序列，有 $Q(4)=20.25(0.09)$，$Q(8)=61.95(0.0004)$，说明方程(9.35)的波动率模型不能充分处理数据的条件异方差性，尤其是在更高延迟时．这并不奇怪，因为单个公共因子仅仅解释了数据广义方差的大约 82.5%．

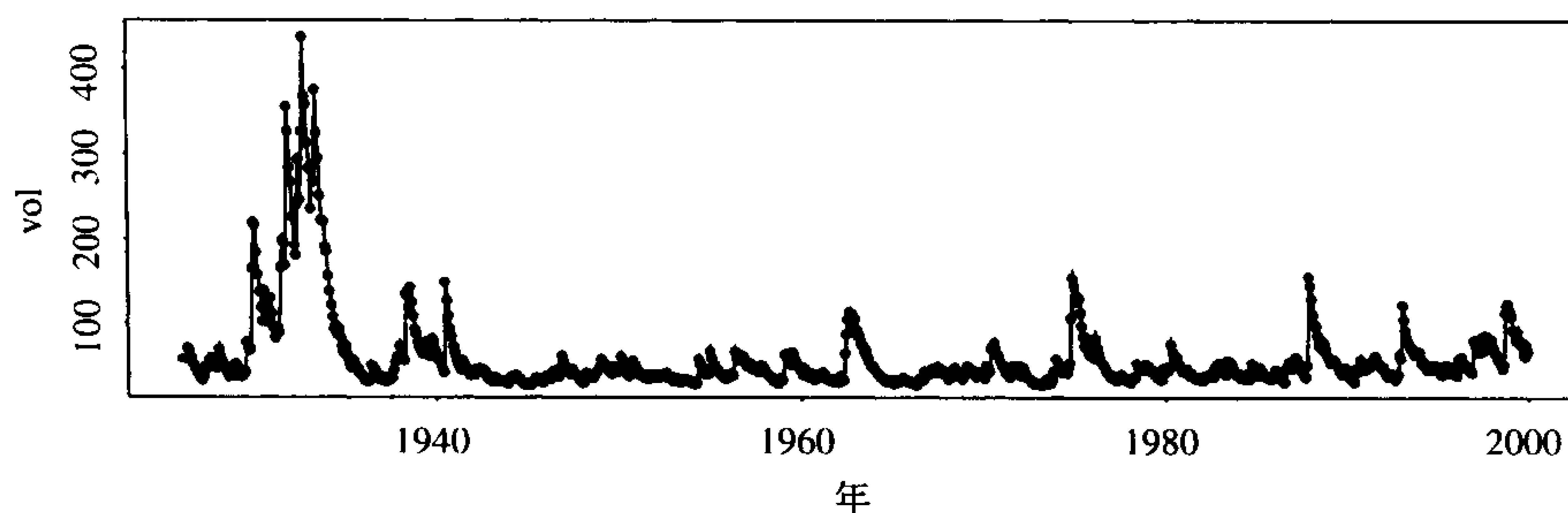
比较方程(9.35)和方程(9.36)中的因子模型与方程(9.24)和方程(9.25)中的时变相关模型，我们看出 a)两个模型的相关方程基本相同，b)如所料想，在波动率方程中，因子模型利

用了较少的参数，c)公共因子模型提供了数据波动率过程的一个合理的近似.

注释：例 9.4 中，我们利用了一个两步估计程序. 第一步中，对公共因子建立了一个波动率模型. 第二步中把波动率的估计当作已给定的来估计多元波动率模型. 这样的估计程序很简单，但是可能并不有效. 更加有效的估计程序是进行联合估计. 这种操作在公共因子已知的条件下相对比较容易. 例如，对例 9.4 中的月对数收益率，如果认为公共因子 $x_t=0.769r_{1t}+0.605r_{2t}$ 是给定的，则可以对方程(9.34)～(9.36)进行联合估计.



a) IBM股票和标准普尔500指数的月对数收益率的第一主成分的时间图



b) 基于GARCH(1, 1)模型拟合的波动率过程

图 9-11

9.5 应用

我们通过考虑多种资产的一个金融头寸的 VaR 来说明多元波动率模型的应用. 假定投资者持有价值各为 100 万美元的 Cisco 系统和 Intel 公司股票的一个多头头寸. 我们利用两只股票从 1991 年 1 月 2 日至 1999 年 12 月 31 日的日对数收益率建立波动率模型. 在数据期间的最后，利用 1 步向前预测和 5％的临界值来计算 VaR.

384～385

令 VaR_1 表示持有 Cisco 系统股票头寸的 VaR，VaR_2 表示持有 Intel 股票头寸的 VaR. 第 7 章中的结果证明了投资者总的日 VaR 为

$$\mathrm{VaR}=\sqrt{\mathrm{VaR}_1^2+\mathrm{VaR}_2^2+2\rho\mathrm{VaR}_1\mathrm{VaR}_2}.$$

在这个解释中，我们考虑计算 VaR 的波动率模型的三种方法．为了简便，我们没有对涉及的参数或模型检验统计量报告其标准误差．所有的参数估计又都在 5%水平下是统计上显著的，而且根据标准化的残差序列及其平方序列可知该模型的 L-B 统计量是充分的．对数收益是以百分比表示的，以致 VaR 计算中的分位数除了 100．令 r_{1t}表示 Cisco 系统股票的收益率，r_{2t}表示 Intel 股票的收益率．

(A)一元模型

这个方法对每只股票收益率都使用一元波动率模型，并用股票收益率的样本相关系数估计 ρ．两只股票收益率的一元波动率模型为

$$r_{1t}=0.380+0.034r_{1,t-1}-0.061r_{1,t-2}-0.055r_{1,t-3}+a_{1t},$$

$$\sigma_{1t}^2=0.599+0.117a_{1,t-1}^2+0.814\sigma_{1,t-1}^2,$$

$$r_{2t}=0.187+a_{2t},$$

$$\sigma_{2t}^2=0.310+0.032a_{2,t-1}^2+0.918\sigma_{2,t-1}^2.$$

样本相关系数为 0.473．VaR 计算中需要的在预测原点 $t=2\,275$ 处的 1 步向前预测为

$$\hat{r}_1=0.626,\quad \hat{\sigma}_1^2=4.152,\quad \hat{r}_2=0.187,\quad \hat{\sigma}_2^2=6.087,\quad \hat{\rho}=0.473.$$

两个日收益率的 5%分位数为

$$q_1=0.626-1.65\sqrt{4.152}=-2.736,\quad q_2=0.187-1.65\sqrt{6.087}=-3.884.$$

其中负号表示损失．单只股票的 VaR 为

$$\mathrm{VaR}_1=1\,000\,000q_1/100=27\,360(\text{美元}),\mathrm{VaR}_2=1\,000\,000q_2/100=38\,840(\text{美元})$$

因此，投资者总的 VaR 为 VaR=57 117 美元．

(B)常相关二元模型

这个方法对股票收益率采用一个二元 GARCH(1，1)模型．假定相关系数是随时间固定不变的，但是它可以与其他参数联合估计．模型为

$$r_{1t}=0.385+0.038r_{1,t-1}-0.060r_{1,t-2}-0.047r_{1,t-3}+a_{1t},$$

$$r_{2t}=0.222+a_{2t},$$

$$\sigma_{11,t}=0.624+0.110a_{1,t-1}^2+0.816\sigma_{11,t-1},$$

$$\sigma_{22,t}=0.664+0.038a_{2,t-1}^2+0.853\sigma_{22,t-1},$$

$$\hat{\rho}=0.475.$$

386

这是一个对角的二元 GARCH(1，1)模型．VaR 计算中在预测原点 $t=2\,275$ 的 1 步向前预测为 $\hat{r}_1=0.373$，$\hat{\sigma}_1^2=4.287$，$\hat{r}_2=0.222$，$\hat{\sigma}_2^2=5.706$，$\hat{\rho}=0.475$．因此，我们有 $\mathrm{VaR}_1=30\,432$ 美元，$\mathrm{VaR}_2=37\,195$ 美元．投资者总的 5%的 VaR 为 VaR=58 180 美元．

(C)时变相关模型

最后，通过楚列斯基分解我们允许相关系数随时间变化．拟合的模型为

$$r_{1t}=0.355+0.039r_{1,t-1}-0.057r_{1,t-2}-0.038r_{1,t-3}+a_{1t},$$

$$
\begin{aligned}
r_{2t} &= 0.206 + a_{2t}, \\
g_{11,t} &= 0.420 + 0.091 b_{1,t-1}^2 + 0.858 g_{11,t-1}, \\
q_{21,t} &= 0.123 + 0.689 q_{21,t-1} - 0.014 a_{2,t-1}, \\
g_{22,t} &= 0.080 + 0.013 b_{2,t-1}^2 + 0.971 g_{22,t-1}.
\end{aligned}
$$

其中 $b_{1t}=a_{1t}$，$b_{2t}=a_{2t}-q_{21,t}a_{1t}$. VaR 计算所需的在预测原点 $t=2\,275$ 处的 1 步向前预测为

$$
\hat{r}_1 = 0.352, \quad \hat{r}_2 = 0.206, \quad \hat{g}_{11} = 4.252, \quad \hat{q}_{21} = 0.421, \quad \hat{g}_{22} = 5.594
$$

因此，我们有$\hat{\sigma}_1^2=4.252$，$\hat{\sigma}_{21}=1.791$，$\hat{\sigma}_2^2=6.348$. 相关系数为$\hat{\rho}=0.345$. 利用这些预测，我们有 $\mathrm{VaR}_1=30\,504$ 美元，$\mathrm{VaR}_2=39\,512$ 美元. 总风险值 VaR=57 648 美元.

三个方法的 VaR 估计是类似的. 一元模型给出了最低的 VaR，而常相关模型产生了最高的 VaR. 差别大约为 1 100 美元. 时变波动率模型是两个极端模型的中和.

9.6 多元 t 分布

实证分析显示，前几节中使用的多元高斯新息可能不能刻画资产收益率的峰度. 在这种情形下，多元学生-t 分布可能是有用的. 有许多形式的多元学生-t 分布，这里对波动率的建模，我们给出一个简单形式.

387 称 k 维随机向量 $\boldsymbol{x}=(x_1, \cdots, x_k)'$服从自由度为 v、参数为$\boldsymbol{\mu}=\boldsymbol{0}$，$\Sigma=\boldsymbol{I}$(单位矩阵)的多元学生-$t$ 分布，如果其概率密度函数(pdf)为

$$
f(\boldsymbol{x} \mid v) = \frac{\Gamma((v+k)/2)}{(\pi v)^{k/2}\Gamma(v/2)}(1+v^{-1}\boldsymbol{x}'\boldsymbol{x})^{-(v+k)/2}, \tag{9.37}
$$

其中 $\Gamma(y)$是伽马函数；见 Mardia，Kent 和 Bibby(1979，第 57 页). 方程(9.37)的每个分量 x_i 的方差为 $v/(v-2)$，从而定义$\boldsymbol{\epsilon}_t=\sqrt{(v-2)/v}\boldsymbol{x}$ 为自由度为 v 的标准化的多元学生-t 分布. 通过变换，$\boldsymbol{\epsilon}_t$ 的概率密度函数是

$$
f(\boldsymbol{\epsilon}_t \mid v) = \frac{\Gamma((v+k)/2)}{[\pi(v-2)]^{k/2}\Gamma(v/2)}[1+(v-2)^{-1}\boldsymbol{\epsilon}_t'\boldsymbol{\epsilon}_t]^{-(v+k)/2}. \tag{9.38}
$$

对波动率模型，记 $\boldsymbol{a}_t=\Sigma_t^{1/2}\boldsymbol{\epsilon}_t$，并假定$\boldsymbol{\epsilon}_t$ 服从方程(9.38)中的多元学生-t 分布. 通过变换，$\boldsymbol{a}_t$ 的概率密度函数是

$$
f(\boldsymbol{a}_t \mid v, \Sigma_t) = \frac{\Gamma((v+k)/2)}{[\pi(v-2)]^{k/2}\Gamma(v/2)\mid \Sigma_t \mid^{1/2}}[1+(v-2)^{-1}\boldsymbol{a}_t'\Sigma_t^{-1}\boldsymbol{a}_t]^{-(v+k)/2}.
$$

而且，如果我们利用 Σ_t 的楚列斯基分解，则变换后的抖动 $\boldsymbol{b}_t$ 的概率密度函数变为

$$
f(\boldsymbol{b}_t \mid v, \boldsymbol{L}_t, \boldsymbol{G}_t) = \frac{\Gamma((v+k)/2)}{[\pi(v-2)]^{k/2}\Gamma(v/2)\prod_{j=1}^{k} g_{jj,t}^{1/2}}\left[1+(v-2)^{-1}\sum_{j=1}^{k}\frac{b_{jt}^2}{g_{jj,t}}\right]^{-(v+k)/2},
$$

其中 $\boldsymbol{a}_t=\boldsymbol{L}_t\boldsymbol{b}_t$，$g_{jj,t}$是 b_{jt}的条件方差. 因为这个概率密度函数并不包含任何矩阵的逆，所以很容易计算数据的条件似然函数.

附录 A　对估计的一些注释

本章中多元 ARMA 模型的估计是利用“科学计算帮手(Scientific Computing Associates)”中的时间序列程序 SCA 进行的．多元波动率模型的估计是利用时间序列的回归分析(RATS)程序进行的．下面是利用 RATS 程序估计多元波动率模型的一些运行流程．以“ * ”开始的行表示“注解”．

(A)：例 9.2 中的对角常相关 AR(2)-GARCH(1，1)模型的估计．该程序包括对每一个分量的一些 L-B 统计量以及最后几个观测的一些拟合值．数据文件为“m-ibmspln. dat”，有两列共 888 个观测． 388

```
all 0 888:1
open data m-ibmspln.dat
data(org=obs) / r1 r2
set h1 = 0.0
set h2 = 0.0
nonlin a0 a1 b1 a00 a11 b11 rho c1 c2 p1
frml a1t = r1(t)-c1-p1*r2(t-1)
frml a2t = r2(t)-c2
frml gvar1 = a0+a1*a1t(t-1)**2+b1*h1(t-1)
frml gvar2 = a00+a11*a2t(t-1)**2+b11*h2(t-1)
frml gdet = -0.5*(log(h1(t)=gvar1(t))+log(h2(t)=gvar2(t)) $
 +log(1.0-rho**2))
frml gln = gdet(t)-0.5/(1.0-rho**2)*((a1t(t)**2/h1(t)) $
 +(a2t(t)**2/h2(t))-2*rho*a1t(t)*a2t(t)/sqrt(h1(t)*h2(t)))
smpl 3 888
compute c1 = 1.22, c2 = 0.57, p1 = 0.1, rho = 0.1
compute a0 = 3.27, a1 = 0.1, b1 = 0.6
compute a00 = 1.17, a11 = 0.13, b11 = 0.8
maximize(method=bhhh,recursive,iterations=150) gln
set fv1 = gvar1(t)
set resi1 = a1t(t)/sqrt(fv1(t))
set residsq = resi1(t)*resi1(t)
* Checking standardized residuals *
cor(qstats,number=12,span=4) resi1
* Checking squared standardized residuals *
cor(qstats,number=12,span=4) residsq
set fv2 = gvar2(t)
set resi2 = a2t(t)/sqrt(fv2(t))
set residsq = resi2(t)*resi2(t)
* Checking standardized residuals *
cor(qstats,number=12,span=4) resi2
* Checking squared standardized residuals *
cor(qstats,number=12,span=4) residsq
* Last few observations needed for computing forecasts *
set shock1 = a1t(t)
set shock2 = a2t(t)
print 885 888 shock1 shock2 fv1 fv2
```

(B)：例 9.2 中的时变相关系数模型的估计

```
all 0 888:1
open data m-ibmspln.dat
data(org=obs) / r1 r2
set h1 = 45.0
set h2 = 31.0
set rho = 0.8
nonlin a0 a1 b1 f1 a00 a11 b11 d11 f11 c1 c2 p1 p3 q0 q1 q2
frml a1t = r1(t)-c1-p1*r1(t-1)-p3*r2(t-2)
frml a2t = r2(t)-c2
frml gvar1 = a0+a1*a1t(t-1)**2+b1*h1(t-1)+f1*h2(t-1)
frml gvar2 = a00+a11*a2t(t-1)**2+b11*h2(t-1)+f11*h1(t-1) $
 +d11*a1t(t-1)**2
frml rh1 = q0 + q1*rho(t-1) $
 + q2*a1t(t-1)*a2t(t-1)/sqrt(h1(t-1)*h2(t-1))
frml rh = exp(rh1(t))/(1+exp(rh1(t)))
frml gdet = -0.5*(log(h1(t)=gvar1(t))+log(h2(t)=gvar2(t)) $
 +log(1.0-(rho(t)=rh(t))**2))
frml gln = gdet(t)-0.5/(1.0-rho(t)**2)*((a1t(t)**2/h1(t)) $
 +(a2t(t)**2/h2(t))-2*rho(t)*a1t(t)*a2t(t)/sqrt(h1(t)*h2(t)))
smpl 4 888
compute c1 = 1.4, c2 = 0.7, p1 = 0.1, p3 = -0.1
compute a0 = 2.95, a1 = 0.08, b1 = 0.87, f1 = -.03
compute a00 = 2.05, a11 = 0.05
compute b11 = 0.92, f11=-.06, d11=.04, q0 = -2.0
compute q1 = 3.0, q2 = 0.1
nlpar(criterion=value,cvcrit=0.00001)
maximize(method=bhhh,recursive,iterations=150) gln
set fv1 = gvar1(t)
set resi1 = a1t(t)/sqrt(fv1(t))
set residsq = resi1(t)*resi1(t)
* Checking standardized residuals *
cor(qstats,number=16,span=4) resi1
* Checking squared standardized residuals *
cor(qstats,number=16,span=4) residsq
set fv2 = gvar2(t)
set resi2 = a2t(t)/sqrt(fv2(t))
set residsq = resi2(t)*resi2(t)
* Checking standardized residuals *
cor(qstats,number=16,span=4) resi2
* Checking squared standardized residuals *
cor(qstats,number=16,span=4) residsq
* Last few observations needed for computing forecasts *
set rhohat = rho(t)
set shock1 = a1t(t)
set shock2 = a2t(t)
print 885 888 shock1 shock2 fv1 fv2 rhohat
```

(C)：利用楚列斯基分解进行例 9.2 中的时变相关系数模型的估计

```
all 0 888:1
open data m-ibmspln.dat
data(org=obs) / r1 r2
set h1 = 45.0
set h2 = 20.0
set q = 0.8
nonlin a0 a1 b1 a00 a11 b11 d11 f11 c1 c2 p1 p3 t0 t1 t2
frml a1t = r1(t)-c1-p1*r1(t-1)-p3*r2(t-2)
```

```
frml a2t = r2(t)-c2
frml v1 = a0+a1*a1t(t-1)**2+b1*h1(t-1)
frml qt = t0 + t1*q(t-1) + t2*a2t(t-1)
frml bt = a2t(t) - (q(t)=qt(t))*a1t(t)
frml v2 = a00+a11*bt(t-1)**2+b11*h2(t-1)+f11*h1(t-1) $
 +d11*a1t(t-1)**2
frml gdet = -0.5*(log(h1(t) = v1(t))+ log(h2(t)=v2(t)))
frml garchln = gdet-0.5*(a1t(t)**2/h1(t)+bt(t)**2/h2(t))
smpl 5 888
compute c1 = 1.4, c2 = 0.7, p1 = 0.1, p3 = -0.1
compute a0 = 1.0, a1 = 0.08, b1 = 0.87
compute a00 = 2.0, a11 = 0.05, b11 = 0.8
compute d11=.04, f11=-.06, t0 =0.2, t1 = 0.1, t2 = 0.1
nlpar(criterion=value,cvcrit=0.00001)
maximize(method=bhhh,recursive,iterations=150) garchln
set fv1 = v1(t)
set resi1 = a1t(t)/sqrt(fv1(t))
set residsq = resi1(t)*resi1(t)
* Checking standardized residuals *
cor(qstats,number=16,span=4) resi1
* Checking squared standardized residuals *
cor(qstats,number=16,span=4) residsq
set fv2 = v2(t)+qt(t)**2*v1(t)
set resi2 = a2t(t)/sqrt(fv2(t))
set residsq = resi2(t)*resi2(t)
* Checking standardized residuals *
cor(qstats,number=16,span=4) resi2
* Checking squared standardized residuals *
cor(qstats,number=16,span=4) residsq
* Last few observations needed for forecasts *
set rhohat = qt(t)*sqrt(v1(t)/fv2(t))
set shock1 = a1t(t)
set shock2 = a2t(t)
set g22 = v2(t)
set q21 = qt(t)
set b2t = bt(t)
print 885 888 shock1 shock2 fv1 fv2 rhohat g22 q21 b2t
```

390

(D)：利用楚列斯基分解进行例 9.3 中的三维时变相关系数波动率模型的估计. 初始的估计值是由序贯建模方法给出.

```
all 0 2275:1
open data d-cscointc.dat
data(org=obs) / r1 r2 r3
set h1 = 1.0
set h2 = 4.0
set h3 = 3.0
set q21 = 0.8
set q31 = 0.3
set q32 = 0.3
nonlin c1 c2 c3 p3 p21 p22 p31 p33 a0 a1 a2 t0 t1 t2 b0 b1 $
       b2 u0 u1 u2 w0 w1 w2 d0 d1 d2 d5
frml a1t = r1(t)-c1-p3*r1(t-3)
frml a2t = r2(t)-c2-p21*r1(t-2)-p22*r2(t-2)
frml a3t = r3(t)-c3-p31*r1(t-1)-p33*r3(t-1)
frml v1 = a0+a1*a1t(t-1)**2+a2*h1(t-1)
frml q1t = t0 + t1*q21(t-1) + t2*a2t(t-1)
```

```
frml bt = a2t(t) - (q21(t)=q1t(t))*a1t(t)
frml v2 = b0+b1*bt(t-1)**2+b2*h2(t-1)
frml q2t = u0 + u1*q31(t-1) + u2*a3t(t-1)
frml q3t = w0 + w1*q32(t-1) + w2*a2t(t-1)
frml b1t = a3t(t)-(q31(t)=q2t(t))*a1t(t)-(q32(t)=q3t(t))*bt(t)
frml v3 = d0+d1*b1t(t-1)**2+d2*h3(t-1)+d5*h2(t-1)
frml gdet = -0.5*(log(h1(t) = v1(t))+ log(h2(t)=v2(t)) $
            +log(h3(t)=v3(t)))
frml garchln = gdet-0.5*(a1t(t)**2/h1(t)+bt(t)**2/h2(t) $
            +b1t(t)**2/h3(t))
smpl 8 2275
compute c1 = 0.07, c2 = 0.33, c3 = 0.19, p1 = 0.1, p3 = -0.04
compute p21 =0.2, p22 = -0.1, p31 = -0.26, p33 = 0.06
compute a0 = .01, a1 = 0.05, a2 = 0.94
compute t0 = 0.28, t1 =0.82, t2 = -0.035
compute b0 = .17, b1 = 0.08, b2 = 0.89
compute u0= 0.04, u1 = 0.97, u2 = 0.01
compute w0 =0.006, w1=0.98, w2=0.004
compute d0 =1.38, d1 = 0.06, d2 = 0.64, d5 = -0.027
nlpar(criterion=value,cvcrit=0.00001)
maximize(method=bhhh,recursive,iterations=250) garchln
set fv1 = v1(t)
set resi1 = a1t(t)/sqrt(fv1(t))
set residsq = resi1(t)*resi1(t)
* Checking standardized residuals *
cor(qstats,number=12,span=4) resi1
* Checking squared standardized residuals *
cor(qstats,number=12,span=4) residsq
set fv2 = v2(t)+q1t(t)**2*v1(t)
set resi2 = a2t(t)/sqrt(fv2(t))
set residsq = resi2(t)*resi2(t)
* Checking standardized residuals *
cor(qstats,number=12,span=4) resi2
* Checking squared standardized residuals *
cor(qstats,number=12,span=4) residsq
set fv3 = v3(t)+q2t(t)**2*v1(t)+q3t(t)**2*v2(t)
set resi3 = a3t(t)/sqrt(fv3(t))
set residsq = resi3(t)*resi3(t)
* Checking standardized residuals *
cor(qstats,number=12,span=4) resi3
* Checking squared standardized residuals *
cor(qstats,number=12,span=4) residsq
* print standardized residuals and correlation-coefficients
set rho21 = q1t(t)*sqrt(v1(t)/fv2(t))
set rho31 = q2t(t)*sqrt(v1(t)/fv3(t))
set rho32 = (q2t(t)*q1t(t)*v1(t) $
 +q3t(t)*v2(t))/sqrt(fv2(t)*fv3(t))
print 10 2275 resi1 resi2 resi3
print 10 2275 rho21 rho31 rho32
print 10 2275 fv1 fv2 fv3
```

391

练习题

1. 文件“m-spibmge.dat”包含了标准普尔500指数、IBM股票、GE股票从1926年1月至1999年12月的月对数收益率，以百分比表示．收益率包括了分红．

392

(a)计算三个收益率序列的样本均值、样本协方差矩阵和样本相关矩阵.

(b)计算三个序列的 1 步延迟和 2 步延迟的交叉相关矩阵. 对于三个序列之间的线性关系做出判断. 有引导-延迟关系吗?

(c)用多元 L-B 统计量在 5%显著性水平下，检验原假设 $H_0: \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_4 = \mathbf{0}$. 得出你的结论.

2. 讨论以百分比表示的 GE 股票和标准普尔 500 指数的月对数收益率. 对这个二元序列建立一个常相关 GARCH 模型. 检验拟合模型的合适性，并得到协方差矩阵在预测原点 1999 年 12 月的 1 步向前预测.
3. 讨论以百分比表示的 GE 股票和标准普尔 500 指数的月对数收益率. 对这个二元序列建立一个时变相关的 GARCH 模型，对相关系数利用 logistic 函数. 检验拟合模型的合适性，并得到协方差矩阵在预测原点 1999 年 12 月的 1 步向前预测.
4. 讨论以百分比表示的 GE 股票和标准普尔 500 指数的月对数收益率. 对这个二元序列利用楚列斯基分解建立一个时变相关的 GARCH 模型. 检验拟合模型的合适性，并得到协方差矩阵在预测原点 1999 年 12 月的 1 步向前预测. 比较所得模型与前面两个问题中所得的模型.
5. 联合考虑三维收益率序列. 利用楚列斯基分解给数据建立多元时变相关波动率模型. 讨论模型的涵义，并计算预测原点为 $t=888$ 的向前 1 步的波动率预测值.
6. 一位投资者想知道他持有 50 万美元的 Dell 股票和 100 万 Cisco Systems 股票的多头头寸的风险值(VaR). 用从 1990 年 2 月 20 日到 1999 年 12 月 31 日的日对数收益率和 5%的临界值来计算. 数据在文件“d-dellcsco9099. dat”. 运用 9.5 节介绍的波动率建模的三种方法，并对结果进行比较.

参考文献

Bollerslev, T. (1990), "Modeling the coherence in short-term nominal exchange rates: A multivariate generalized ARCH approach," *Review of Economics and Statistics*, 72, 498–505.

Chib, S., Nardari, F., and Shephard, N. (1999), "Analysis of high dimensional multivariate stochastic volatility models," Working paper, Washington University, St Louis.

Harvey, A., Ruiz, E., and Shephard, N. (1995), "Multivariate stochastic variance models," in *ARCH Selected Readings*, ed. R. F. Engle, pp. 253–276, Oxford University Press: Oxford, UK.

393

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, Academic Press: New York.

Pourahmadi, M. (1999), "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterization," *Biometrika*, 86, 677–690.

Tse, Y. K. (2000), "A test for constant correlations in a multivariate GARCH model," *Journal of Econometrics*, 98, 107–127.

Tse, Y. K., and Tsui, A. K. C. (1998), "A multivariate GARCH model with time-varying correlations," working paper, Department of Economics, National University of Singapore.

394

第 10 章　马尔可夫链蒙特卡罗方法的应用

计算设施与计算方法的进步已经大大地增强了我们解决复杂问题的能力．这种进步也扩展了许多已有的计量经济和统计方法的应用．统计中这种成果的例子包括马尔可夫链蒙特卡罗(MCMC)方法和数据扩张．这些方法使得我们能够进行一些几年前还不可行的统计推断．本章中，我们引进金融中广泛使用的 MCMC 方法和数据扩张的思想．特别地，我们讨论经由吉布斯抽样的贝叶斯推断，并描述 MCMC 方法的各种应用．MCMC 方法论的迅速发展使涵盖所有在文献中出现的新方法是不可能的．感兴趣的读者可以参考一些近期的关于贝叶斯统计和经验贝叶斯统计的书(例如 Carlin 和 Louis，2000；Gelman，Carlin，Stern 和 Rubin，1995)．

对于应用，我们把注意力放在与金融计量经济学有关的问题上．本章中所表述的仅仅是这个方法在金融中所有可能应用的一小部分．事实上，可以说，我们这里讨论的贝叶斯推断和 MCMC 方法在金融计量经济学的大多数研究(如果不是全部的话)中，都是切实可行的．

我们先回顾一下马尔可夫过程的概念．考虑一个随机过程$\{X_t\}$，这里假定每个 X_t 都在空间$\boldsymbol{\Theta}$上取值．称过程$\{X_t\}$是一个马尔可夫过程，如果它具有性质：给定 X_t 的值，$X_h(h>t)$的值不依赖于 $X_s(s<t)$的取值．换句话说，称$\{X_t\}$是一个马尔可夫过程，如果它的条件分布函数满足

$$P(X_h \mid X_s, s \leqslant t) = P(X_h \mid X_t), \qquad h > t.$$

如果$\{X_t\}$是一个离散时间的随机过程，则前面的性质变为

395

$$P(X_h \mid X_t, X_{t-1}, \cdots) = P(X_h \mid X_t), \qquad h > t.$$

令 $\boldsymbol{A}$ 表示$\boldsymbol{\Theta}$的子集，则函数

$$P_t(\boldsymbol{\theta}, h, \boldsymbol{A}) = P(X_h \in \boldsymbol{A} \mid X_t = \boldsymbol{\theta}), \qquad h > t,$$

称为马尔可夫过程的转移概率函数．如果转移概率依赖于 $h-t$，但是不依赖于 t，则过程有一个平稳的转移分布．

10.1　马尔可夫链模拟

考虑参数向量为 $\boldsymbol{\theta}$ 和数据为 $\boldsymbol{X}$ 的推断问题，这里$\boldsymbol{\theta} \in \boldsymbol{\Theta}$．为了作出推断，我们需要了解分布 $P(\boldsymbol{\theta} \mid \boldsymbol{X})$．马尔可夫链模拟的思想是在$\boldsymbol{\Theta}$上模拟一个马尔可夫过程，它收敛于平稳转移分布 $P(\boldsymbol{\theta} \mid \boldsymbol{X})$．

马尔可夫链模拟的关键是构造一个具有指定的平稳转移分布 $P(\boldsymbol{\theta} \mid \boldsymbol{X})$的马尔可夫过程，并且充分长地运行这个模拟，使得过程当前值的分布与平稳转移分布足够接近．对给定的 $P(\boldsymbol{\theta} \mid \boldsymbol{X})$，可以证明能够构造许多具有所需性质的马尔可夫链．我们将利用马尔可夫链模拟来

得到分布 $P(\boldsymbol{\theta} \mid \boldsymbol{X})$的方法称为 MCMC 方法.

统计文献中，MCMC 方法以各种各样的形式发展. 考虑数据分析中“缺失值”的问题. 本书中讨论的大多数统计方法都是在“完全数据”(即没有缺失值)的假定下发展的. 例如，在模拟资产收益率的日波动率时，我们假定样本期间的收益率数据在除了周末和假期的所有交易日都是可以得到的. 如果有一个缺失值，我们应该怎么办?

Dempster，Laird 和 Rubin(1977)提出一个称为 EM 算法的迭代方法来解决这个问题. 这个方法包含两个步骤. 第一个步骤，如果缺失值是可以得到的，则我们能够利用完全数据分析的方法来建立一个波动率模型. 第二，给定可以利用的数据及拟合的模型，我们能够推导出缺失值的统计分布. 填充缺失值的一个简单方法是利用缺失值的导出分布的条件期望. 实际中，可以用缺失值的一个任意值来开始这个方法，迭代足够多的次数直到收敛. 前面程序中的第一个步骤涉及对一个指定的模型实施最大似然估计，并称为 M 步. 第二个步骤是计算缺失值的条件期望，称为 E 步.

Tanner 和 Wong(1987)以两种方式扩展了 EM 算法. 首先，他们引进了迭代模拟的思想. 例如，可以简单地利用从缺失值的导出条件分布里面抽取的一个随机数来代替缺失值，而不是用条件期望来代替. 第二，他们利用数据扩张的概念扩展了 EM 算法的应用. 数据扩张的意思是在研究的问题中加入一个辅助变量. 可以证明许多的模拟方法通常可以通过数据扩张来简单化或加速，见本章中的应用一节. [396]

10.2 吉布斯抽样

Geman 和 Geman(1984)、Gelfand 和 Smith(1990)的吉布斯抽样(或吉布斯样本)可能是最流行的 MCMC 方法. 我们通过使用一个三个参数的简单问题来引进吉布斯抽样的思想. 这里所用的单词“参数”意义非常广泛. 一个缺失的数据点可以认为是 MCMC 框架下的一个参数. 类似地，当有 N 个交易价格可以得到时，则一个不可观测的变量如资产的“真实”价格可以认为是 N 个参数. “参数”这个概念与数据扩张是相关的，当我们讨论 MCMC 方法的应用时就变得很清楚.

将三个参数表示为 θ_1、θ_2 和 θ_3. 令 $\boldsymbol{X}$ 表示可用的数据集，$\boldsymbol{M}$ 表示接受的模型. 这里的目标是估计这些参数，以便利用拟合的模型作出推断. 假定模型的似然函数很难得到，但是在给定其他参数下，单个参数的三个条件分布是可以得到的. 换句话说，我们假定已知下面的三个条件分布：

$$f_1(\theta_1 \mid \theta_2,\theta_3,\boldsymbol{X},\boldsymbol{M}),\quad f_2(\theta_2 \mid \theta_3,\theta_1,\boldsymbol{X},\boldsymbol{M}),\quad f_3(\theta_3 \mid \theta_1,\theta_2,\boldsymbol{X},\boldsymbol{M}),\qquad (10.1)$$

其中 $f_i(\theta_i \mid \theta_{j\neq i},\boldsymbol{X},\boldsymbol{M})$表示给定数据、模型以及其他两个参数的条件下，参数 θ_i 的条件分布. 应用中，我们不需要知道条件分布的精确形式. 需要的是从三个条件分布的每一个中抽取随机数的能力.

令 $\theta_{2,0}$和 $\theta_{3,0}$是 θ_2 和 θ_3 的两个任意初始值，则吉布斯抽样过程进行如下：

1)从 $f_1(\theta_1 \mid \theta_{2,0},\theta_{3,0},\boldsymbol{X},\boldsymbol{M})$中抽取一个随机样本，并将抽取的随机数表示为 $\theta_{1,1}$.

2)从 $f_2(\theta_2 \mid \theta_{3,0}, \theta_{1,1}, \boldsymbol{X}, \boldsymbol{M})$ 中抽取一个随机样本，并将抽取的随机数表示为 $\theta_{2,1}$.

3)从 $f_3(\theta_3 \mid \theta_{1,1}, \theta_{2,1}, \boldsymbol{X}, \boldsymbol{M})$ 中抽取一个随机样本，并将抽取的随机数表示为 $\theta_{3,1}$. 这就完成了一个吉布斯迭代，且参数变为 $\theta_{1,1}$、$\theta_{2,1}$ 和 $\theta_{3,1}$.

下一步，利用新参数作为初始值，重复前面随机抽取的迭代，我们可以完成另一个吉布斯迭代，得到更新的参数 $\theta_{1,2}$、$\theta_{2,2}$ 和 $\theta_{3,2}$. 我们可以重复前面的迭代 m 次，得到一系列的随机抽取：

397

$$(\theta_{1,1}, \theta_{2,1}, \theta_{3,1}), \cdots, (\theta_{1,m}, \theta_{2,m}, \theta_{3,m}).$$

在一些正则条件下，可以证明对一个充分大的 m，$(\theta_{1,m}, \theta_{2,m}, \theta_{3,m})$ 渐近等价于来自于三个参数的联合分布 $f(\theta_1, \theta_2, \theta_3 \mid \boldsymbol{X}, \boldsymbol{M})$ 的一个随机抽取. 正则条件是弱的；本质上是要求对一个任意的初始值 $(\theta_{1,0}, \theta_{2,0}, \theta_{3,0})$，前述的吉布斯迭代有机会访问整个参数空间. 真实的收敛定理涉及到利用马尔可夫链的理论；见 Tierney(1994).

实际中，我们利用一个充分大的 n，并且丢掉吉布斯迭代的前 m 个随机抽取，这样建立一个吉布斯样本，即

$$(\theta_{1,m+1}, \theta_{2,m+1}, \theta_{3,m+1}), \cdots, (\theta_{1,n}, \theta_{2,n}, \theta_{3,n}). \tag{10.2}$$

因为前面的迭代从联合分布 $f(\theta_1, \theta_2, \theta_3 \mid \boldsymbol{X}, \boldsymbol{M})$ 中建立了一个随机样本，所以可以利用它们来作出统计推断. 例如，θ_i 的点估计与其方差为

$$\hat{\theta}_i = \frac{1}{n-m}\sum_{j=m+1}^{n}\theta_{i,j}, \quad \hat{\sigma}_i^2 = \frac{1}{n-m-1}\sum_{j=m+1}^{n}(\theta_{i,j} - \hat{\theta}_i)^2 \tag{10.3}$$

方程(10.2)中的吉布斯样本可以有许多种使用方法. 例如，如果感兴趣于检验原假设 $H_a: \theta_1 = \theta_2$ 对备择假设 $H_a: \theta_1 \neq \theta_2$，则可以简单地得到 $\theta = \theta_1 - \theta_2$ 的点估计与其方差为

$$\hat{\theta} = \frac{1}{n-m}\sum_{j=m+1}^{n}(\theta_{1,j} - \theta_{2,j}), \quad \hat{\sigma}^2 = \frac{1}{n-m-1}\sum_{j=m+1}^{n}(\theta_{1,j} - \theta_{2,j} - \hat{\theta})^2.$$

则原假设可以利用传统的 t 比率统计量 $t = \hat{\theta}/\hat{\sigma}$ 来检验.

注释：吉布斯抽样中被丢掉的前 m 个随机抽取通常称为 burn-ins 样本. Burn-ins 用来保证方程(10.2)中的吉布斯样本确实与从联合分布 $f(\theta_1, \theta_2, \theta_3 \mid \boldsymbol{X}, \boldsymbol{M})$ 中的随机样本足够接近.

注释：前面讨论的方法包含了运行单个长链并在 burn-ins 后保留所有随机抽取得到一个吉布斯样本. 另一种方法，可以利用不同的初始值和一个相对小的 n 运行许多相对短的链. 则每个链中，最后一个吉布斯迭代的随机抽取都可以用来建立一个吉布斯样本.

由前面的介绍，吉布斯抽样具有将一个高维的估计问题利用所有参数的条件分布分解为几个较低维数问题的优点. 极端地，一个具有 N 个参数的高维问题可以通过利用 N 个 1 维的条件分布迭代地解决. 这个性质使得吉布斯抽样非常简单且广泛应用. 然而，将所有的随机抽取都简化为 1 维问题通常也是不合适的. 当参数高度相关时，值得将它们联合地抽取. 考虑三个参数的例子作为解释，如果 θ_1 和 θ_2 是高度相关的，则可以在任何可能的情况下，采用条件分

398

布 $f(\theta_1, \theta_2 \mid \theta_3, \boldsymbol{X}, \boldsymbol{M})$ 和 $f_3(\theta_3 \mid \theta_1, \theta_2, \boldsymbol{X}, \boldsymbol{M})$，则一个吉布斯迭代包含(a)给定 θ_3 下联合抽取(θ_1, θ_2)，(b)给定(θ_1, θ_2)下抽取 θ_3. 关于参数相关性对吉布斯抽样的收敛速度的影响，更多的信息可以参见 Liu，Wong 和 Kong(1994).

实际中，吉布斯样本的收敛性是一个重要问题. 理论上仅仅指出当迭代次数 m 充分大时收敛发生，没有对 m 的选择提供具体的指导. 文献中已经建议了许多种检验吉布斯样本收敛性的方法. 但是并没有关于哪个方法执行的最好的一致结论. 事实上，没有任何一个可以利用的方法能 100%地保证所研究的吉布斯样本对所有的应用都是收敛的. 检验方法的性能通常依赖于所处理的问题. 实际应用中必须仔细地执行，以保证没有明显的对收敛性要求的违背；对收敛性的检验方法见 Carlin 和 Louis(2000)以及 Gelman 等(1995). 应用中，以不同的初始值重复吉布斯抽样几次，保证算法收敛是很重要的.

10.3 贝叶斯推断

在吉布斯抽样中条件分布起着重要的作用. 在统计文献中，这些条件分布称为条件后验分布，因为它们是在数据、其他参数和一定的模型给定的条件下参数的分布. 这一节我们回顾一些在 MCMC 方法中有用的有名的后验分布.

10.3.1 后验分布

有两种统计推断的方法. 第一种方法是基于最大似然原理的经典方法. 这里模型是通过最大化数据的似然函数估计的，且利用拟合的模型作出推断. 另一种方法是贝叶斯推断，它将先验的思想与数据结合，得到后验分布，然后基于后验分布进行统计推断. 历史上，在两类统计推断之间有着激烈的争论，但两种方法都证明是有用的，而且现在都被广泛接受. 到目前为止，本书讨论的方法都属于经典方法. 然而，对所考虑的所有问题贝叶斯解都是存在的. 近年来，随着 MCMC 方法的发展尤其如此，因为 MCMC 方法大大改善了贝叶斯分析的可行性. 读者可以重新回顾一下前几章，并导出所考虑问题的 MCMC 解. 大多数情形下，贝叶斯解类似于我们以前已有的结果. 有些情形，贝叶斯解可能优于其他解. 例如，考虑第 7 章中 VaR 的计算，可以很容易地在 VaR 计算的不确定性中考虑一个贝叶斯解. 然而，这种方法要求大强度的计算. 399

令$\boldsymbol{\theta}$为所采用模型的未知参数向量，$\boldsymbol{X}$ 是数据. 贝叶斯分析寻求将关于参数的知识与数据相结合来作出推断. 参数的知识是通过对参数预先指定一个先验分布表示的，记为 $P(\boldsymbol{\theta})$. 对一个给定的模型，用 $f(\boldsymbol{X} \mid \boldsymbol{\theta})$表示数据的似然函数. 则由条件概率的定义，

$$f(\boldsymbol{\theta} \mid \boldsymbol{X}) = \frac{f(\boldsymbol{\theta}, \boldsymbol{X})}{f(\boldsymbol{X})} = \frac{f(\boldsymbol{X} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta})}{f(\boldsymbol{X})} \tag{10.4}$$

这里边际分布可以通过下式得到

$$f(\boldsymbol{X}) = \int f(\boldsymbol{X}, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = \int f(\boldsymbol{X} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$$

方程(10.4)中的分布 $f(\boldsymbol{\theta} \mid \boldsymbol{X})$称为$\boldsymbol{\theta}$的后验分布. 一般地，我们可以利用贝叶斯准则得到

$$f(\boldsymbol{\theta} \mid \boldsymbol{X}) \propto f(\boldsymbol{X} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta}) \tag{10.5}$$

其中 $P(\boldsymbol{\theta})$是先验分布，$f(\boldsymbol{X} \mid \boldsymbol{\theta})$是似然函数. 由方程(10.5)，基于似然函数 $f(\boldsymbol{X} \mid \boldsymbol{\theta})$作出统计推断，相当于利用具有一个固定的先验分布的贝叶斯方法.

10.3.2 共轭先验分布

方程(10.4)中得到的后验分布一般不是简单的，但是，也有先验分布与后验分布属于同样的分布族的情形. 这样一个先验分布称为共轭先验分布. 对 MCMC 方法，共轭先验的使用可以得到条件后验分布的一个显式解，然后则可以利用通常的概率分布的计算机程序得到吉布斯样本的随机抽取. 下面，我们回顾一些著名的共轭先验. 更多的信息，读者可以参考有关贝叶斯统计的书(如 DeGroot，1970，第 9 章).

结果 1：假定 x_1，…，x_n 是从均值为μ(未知)、方差为σ^2(已知且为正数)的正态分布中抽取的一个随机样本. 假定 μ 的先验分布是均值为μ_o、方差为 σ_o^2 的正态分布，则给定数据和先验分布下，μ 的后验分布为均值为μ_*、方差为 σ_*^2 的正态分布，μ_* 和 σ_*^2 由下式给出：

400

$$\mu_* = \frac{\sigma^2 \mu_o + n\sigma_o^2 \bar{x}}{\sigma^2 + n\sigma_o^2} \quad 且\ \sigma_*^2 = \frac{\sigma^2 \sigma_o^2}{\sigma^2 + n\sigma_o^2},$$

其中$\bar{x} = \sum_{i=1}^{n} x_i / n$ 是样本均值.

在贝叶斯分析中，通常利用精度参数 $\eta = 1/\sigma^2$(即方差 σ^2 的逆)非常方便. 将先验分布的精度参数表示为 $\eta_o = 1/\sigma_o^2$，后验分布的精度参数表示为 $\eta_* = 1/\sigma_*^2$，则结果 1 可以改写为

$$\eta_* = \eta_o + n\eta \ 且\ \mu_* = \frac{\eta_o}{\eta_*} \times \mu_o + \frac{n\eta}{\eta_*} \times \bar{x}.$$

对所考虑的正态随机样本，关于 μ 的数据信息包含进了样本均值$\bar{x}$，它是 μ 的充分统计量. $\bar{x}$ 的精度为 $n/\sigma^2 = n\eta$，因此，结果 1 说明(a)后验分布的精度为先验分布的精度和数据的精度之和. (b)后验均值是先验均值和样本均值的加权平均，权重与精度成正比. 这两个公式也说明了当样本量 n 增加时，先验分布的贡献降低.

结果 1 的一个多元形式在 MCMC 方法中涉及线性回归模型时特别有用；参考 Box 和 Tiao (1973).

结果 1a：假定 $\boldsymbol{x}_1$，…，$\boldsymbol{x}_n$ 是从均值向量为$\boldsymbol{\mu}$、已知协方差矩阵为 Σ的多元正态分布建立的随机样本. 又假定$\boldsymbol{\mu}$ 的先验分布是均值向量为$\boldsymbol{\mu}_o$、协方差矩阵为 Σ_o的多元正态，则$\boldsymbol{\mu}$ 的后验分布也是多元正态的，其均值向量为$\boldsymbol{\mu}_*$、协方差矩阵为Σ_*，这里

$$\Sigma_*^{-1} = \Sigma_o^{-1} + n\Sigma^{-1}, \quad \boldsymbol{\mu}_* = \Sigma_*(\Sigma_o^{-1}\boldsymbol{\mu}_o + n\Sigma^{-1}\bar{\boldsymbol{x}}),$$

其中 $\bar{\boldsymbol{x}} = \sum_{i=1}^{n} \boldsymbol{x}_i / n$ 是样本均值，它服从均值为$\boldsymbol{\mu}$、协方差矩阵为Σ/n 的多元正态分布. 注意到 $n\Sigma^{-1}$是$\bar{\boldsymbol{x}}$的精度矩阵，且 Σ_o^{-1} 是先验分布的精度矩阵.

称一个随机变量 η 服从具有正参数α 和β 的伽马分布，如果它的分布密度函数为

$$f(\eta \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \eta^{\alpha-1} e^{-\beta\eta}, \qquad \eta > 0,$$

这里 $\Gamma(\alpha)$为伽马函数. 对于这个分布，$E(\eta)=\alpha/\beta$，$\operatorname{Var}(\eta)=\alpha/\beta^2$. 401

结果 2：假定 x_1，…，x_n 是从一个给定均值为μ、未知精度为η的正态分布中抽取的随机样本. 如果η的先验分布是具有正参数α和β的伽马分布，则η的后验分布也是伽马分布，其参数为$\alpha+(n/2)$和$\beta+\sum_{i=1}^{n}(x_i-\mu)^2/2$.

称一个随机变量θ具有正参数α和β的贝塔分布，如果它的概率密度函数为

$$f(\theta\mid\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\qquad 0<\theta<1.$$

θ的均值为$E(\theta)=\alpha/(\alpha+\beta)$，方差为$\operatorname{Var}(\theta)=\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$.

结果 3：假定 x_1，…，x_n 是从参数为θ的伯努利分布中抽取的随机样本. 如果θ的先验分布是具有给定的正参数α和β的贝塔分布，则θ的后验分布也是贝塔分布，参数为$\alpha+\sum_{i=1}^{n}x_i$和$\beta+n-\sum_{i=1}^{n}x_i$.

结果 4：假定 x_1，…，x_n 是从一个参数为λ的泊松分布中抽取的随机样本. 又假定λ的先验分布是具有给定的正参数α和β的伽马分布，则λ的后验分布也是伽马分布，参数为$\alpha+\sum_{i=1}^{n}x_i$和$\beta+n$.

结果 5：假定 x_1，…，x_n 是从一个参数为λ的指数分布中抽取的随机样本. 如果λ的先验分布是具有给定的正参数α和β的伽马分布，则λ的后验分布也是伽马分布，参数为$\alpha+n$和$\beta+\sum_{i=1}^{n}x_i$.

称一个随机变量X具有参数为m和λ的负二项分布，这里$m>0$，$0<\lambda<1$，如果它的概率函数为

$$p(n\mid m,\lambda)=\begin{cases}\dbinom{m+n-1}{n}\lambda^m(1-\lambda)^n & \text{若 } n=0,1,\cdots\\ 0 & \text{其他}\end{cases}$$

金融中负二项分布的一个简单例子是：假定应聘者是独立的，并且每个应聘者是最好人选的概率为λ. 在一个公司发现对它的m个空缺的m个“恰当的候选人”以前，需要面试多少个 MBA 毕业生？用Y表示全部应聘者的数量，则$X=Y-m$服从参数为m和λ的负二项分布.

结果 6：假定 x_1，…，x_n 服从一个参数为m和λ的负二项分布中建立的一个随机样本，这里m是正的，而且是固定的. 如果λ的先验分布是具有正参数α和β的贝塔分布，则λ的后验分布也是贝塔分布，参数为$\alpha+mn$和$\beta+\sum_{i=1}^{n}x_i$. 402

下面我们考虑具有未知均值μ和未知精度η的正态分布的情形. 二维的先验分布分解为$P(\mu,\ \eta)=P(\mu\mid\eta)P(\eta)$.

结果 7：假定 x_1，…，x_n 是从未知均值为μ和未知精度为η的正态分布中抽取的随机样

本. 又假定给定 $\eta=\eta_o$ 下，μ 的条件分布是均值为 μ_o、精度 $\tau_o\eta_o$ 的正态分布，而且 η 的边际分布是具有正参数 α 和 β 的伽马分布，则给定 $\eta=\eta_o$ 下，μ 的条件后验分布是均值为 μ_*、精度 η_* 的正态分布，

$$\mu_* = \frac{\tau_o\mu_o + n\bar{x}}{\tau_o + n} \text{ 且 } \eta_* = (\tau_o + n)\eta_o,$$

其中 $\bar{x} = \sum_{i=1}^{n} x_i/n$ 是样本均值，且 η 的边际后验分布是参数为 $\alpha + (n/2)$ 和 β_* 的伽马分布，这里

$$\beta_* = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\tau_o n(\bar{x} - \mu_o)^2}{2(\tau_o + n)}.$$

当对随机变量的条件方差感兴趣时，通常使用逆 χ^2 分布. 称随机变量 Y 服从自由度为 v 的逆 χ^2 分布，如果 $1/Y$ 服从自由度为 v 的 χ^2 分布. Y 的概率密度函数为

$$f(y \mid v) = \frac{2^{-v/2}}{\Gamma(v/2)} y^{-(v/2+1)} e^{-1/(2y)}, \qquad y > 0.$$

对于这个分布，我们有：如果 $v>2$，则 $E(Y)=1/(v-2)$；如果 $v>4$，则 $\mathrm{Var}(Y)=2/[(v-2)^2(v-4)]$.

结果 8： 假定 $a_1, \cdots, a_n$ 是从一个均值为 0、方差为 σ^2 的正态分布中抽取的随机样本. 又假定 σ^2 的先验分布是自由度为 v 的逆 χ^2 分布(即，$(v\lambda)/\sigma^2 \sim \chi_v^2$，其中 $\lambda>0$)，则 σ^2 的后验分布也是逆 χ^2 分布，其自由度为 $v+n$，即，$\left(v\lambda + \sum_{i=1}^{n} a_i^2\right)/\sigma^2 \sim \chi_{v+n}^2$.

10.4 其他算法

403 许多应用中，对于条件后验分布没有显式解. 但是统计文献中已经提出了许多聪明的算法来克服这个困难. 本节中，我们讨论一些这样的算法.

10.4.1 Metropolis 算法

当条件后验分布已知，并且不是标准化的常数时，这个算法是可行的；参考 Metropolis 和 Ulam(1949)以及 Metropolis 等(1953). 假定我们希望从分布 $f(\boldsymbol{\theta} \mid \boldsymbol{X})$中抽取一个随机样本，它包含一个复杂的标准化常数，直接的抽取要么太浪费时间，要么不可行. 但是存在一个近似分布，可以很容易地得到随机抽取. Metropolis 算法就是从近似分布中产生一系列的随机抽取，其分布函数收敛到 $f(\boldsymbol{\theta} \mid \boldsymbol{X})$. 此算法如下进行：

1)抽取一个随机的初始值 $\boldsymbol{\theta}_o$，满足 $f(\boldsymbol{\theta}_o \mid \boldsymbol{X})>0$.

2)对 $t=1, 2, \cdots$

(a)第 t 次迭代时，在给定前面的抽取 $\boldsymbol{\theta}_{t-1}$ 下，从已知分布中抽取一个候选样本 $\boldsymbol{\theta}_*$. 用 $J_t(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1})$表示已知分布，在 Gelman 等(1995)中称为跳跃分布. 这个跳跃分布一定是对称

的，即对于所有的$\boldsymbol{\theta}_i$，$\boldsymbol{\theta}_j$ 和 t，有 $J_t(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_j)=J_t(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_i)$.

(b)计算比率

$$r=\frac{f(\boldsymbol{\theta}_* \mid \boldsymbol{X})}{f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{X})}$$

(c)设定

$$\boldsymbol{\theta}_t=\begin{cases}\boldsymbol{\theta}_* & \text{以概率 } \min(r,1)\\ \boldsymbol{\theta}_{t-1} & \text{其他}\end{cases}$$

在一些正规性条件下，序列$\{\boldsymbol{\theta}_t\}$依分布收敛到 $f(\boldsymbol{\theta} \mid \boldsymbol{X})$；参见 Gelman 等(1995).

算法的实施要求对所有的$\boldsymbol{\theta}_*$和$\boldsymbol{\theta}_{t-1}$计算比率 r 的能力，以便从跳跃分布中抽取$\boldsymbol{\theta}_*$，并从均匀分布中抽取一个随机实现，决定接受或者拒绝$\boldsymbol{\theta}_*$. 不需要 $f(\boldsymbol{\theta} \mid \boldsymbol{X})$的标准化常数，因为只利用比率.

此算法的接受和拒绝准则可以陈述如下：(ⅰ)如果从$\boldsymbol{\theta}_{t-1}$到$\boldsymbol{\theta}_*$的跳跃增加了条件后验密度，则接受$\boldsymbol{\theta}_*$作为$\boldsymbol{\theta}_t$；(ⅱ)如果这个跳跃降低了后验密度，则以等于密度比 r 的概率设定$\boldsymbol{\theta}_t=\boldsymbol{\theta}_*$；否则设定$\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1}$. 这个程序看上去是合理的.

对称跳跃分布的例子包括均值参数的正态和学生-t 分布. 对一个给定的协方差矩阵，我们有 $f(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_j)=f(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_i)$，其中 $f(\boldsymbol{\theta} \mid \boldsymbol{\theta}_o)$表示均值向量为$\boldsymbol{\theta}_o$ 的多元正态密度函数. [404]

10.4.2 Metropolis-Hasting 算法

Hasting(1970)以两种方式推广了 Metropolis 算法. 首先，跳跃分布没有必要一定是对称的. 第二，跳跃准则修正为

$$r=\frac{f(\boldsymbol{\theta}_* \mid \boldsymbol{X})/J_t(\boldsymbol{\theta}_* \mid \boldsymbol{\theta}_{t-1})}{f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{X})/J_t(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_*)}=\frac{f(\boldsymbol{\theta}_* \mid \boldsymbol{X})J_t(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{\theta}_*)}{f(\boldsymbol{\theta}_{t-1} \mid \boldsymbol{X})J_t(\boldsymbol{\theta}_* \mid \boldsymbol{\theta}_{t-1})}$$

这个修正的算法称为 Metropolis-Hasting 算法.

10.4.3 格子吉布斯抽样

金融应用中，所采用的模型可能包含一些非线性参数(如：ARMA 模型中的滑动平均参数或波动率模型中的 GARCH 参数). 因为非线性参数的条件后验分布没有显式表示，所以，在这种情形下，执行吉布斯抽样可能变得复杂，即使对 Metropolis-Hasting 算法也是如此. Tanner(1996)描述了当条件后验分布是 1 元时，在吉布斯抽样中得到随机抽取的一个简单程序. 这个方法称为格子吉布斯(Griddy Gibbs)抽样，是广泛实用的. 然而，在实际应用中此方法可能并不有效.

令 θ_i 表示具有条件后验分布 $f(\theta_i \mid \boldsymbol{X}, \boldsymbol{\theta}_{-i})$的尺度参数，这里$\boldsymbol{\theta}_{-i}$是剔除 θ_i 之后的参数向量. 例如，如果$\boldsymbol{\theta}=(\theta_1, \theta_2, \theta_3)'$，则$\boldsymbol{\theta}_{-1}=(\theta_2, \theta_3)'$. 格子吉布斯抽样进行如下：

1)从 θ_i 的一个恰当选择的区间上选择格子点，即 $\theta_{i1}\leqslant\theta_{i2}\leqslant\cdots\leqslant\theta_{im}$. 估计条件后验密度函数，得到 $w_j=f(\theta_{ij} \mid \boldsymbol{X}, \boldsymbol{\theta}_{-i})$ ($j=1, \cdots, m$).

2)利用 $w_1, \cdots, w_m$ 得到 $f(\theta_i \mid \boldsymbol{X}, \boldsymbol{\theta}_{-i})$逆累积分布函数的一个近似.

3)抽取一个均匀(0，1)随机变量，并通过近似的逆累积分布函数转换这个观测，得到 θ_i 的一个随机抽取.

下面依次给出关于格子吉布斯抽样的注释. 首先，条件后验分布 $f(\theta_i \mid \boldsymbol{X}, \boldsymbol{\theta}_{-i})$ 的标准化常数并不需要. 因为，逆累积分布函数可以从 $\{w_j\}_{j=1}^m$ 直接得到. 其次，逆累积分布函数的一个简单近似是 $\{\theta_{ij}\}_{j=1}^m$ 离散分布，概率 $p(\theta_{ij}) = w_j / \sum_{v=1}^m w_v$. 第三，在实际应用中，参数 θ_i 的区间 $[\theta_{i1}, \theta_{im}]$ 的选择必须仔细检验. 一个简单的检验程序是考虑 θ_i 的吉布斯抽样的直方图. 如果该直方图显示 θ_{i1} 或 θ_{im} 周围的概率较大，则必须扩展此区间. 然而，如果该直方图显示概率集中在区间 $[\theta_{i1}, \theta_{im}]$ 内，则此区间太宽了，应缩短. 如果区间太宽，因为大多数 w_j 将是 0，所以格子吉布斯抽样变得无效. 最后格子吉布斯或 Metropolis-Hasting 算法可以用在吉布斯抽样中，以得到一些参数的随机抽取.

405

10.5 带时间序列误差的线性回归

我们准备考虑 MCMC 方法的一些具体应用. 下面几节中讨论的例子仅仅是为了解释. 目标是强调这些方法的用途和实用性. 弄懂这些例子有助于读者深入了解 MCMC 方法在金融中的应用.

第一个例子是估计一个带序列相关误差的回归模型. 这是第 2 章中讨论的一个主题，那里我们用 SCA 进行估计. 该模型的一个简单形式为

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + z_t$$
$$z_t = \phi z_{t-1} + a_t$$

其中 y_t 是响应变量；x_{it} 是解释变量；它可能包含 y_t 的延迟值；z_t 服从简单的 AR(1)模型，$\{a_t\}$ 是独立同分布的正态随机变量序列，均值为 0，方差 σ^2. 用 $\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi, \sigma^2)'$ 表示这个模型的参数，其中 $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)'$，令 $\boldsymbol{x}_t = (1, x_{1t}, \cdots, x_{kt})'$ 为时刻 t 时所有回归因子的向量，包括一个单位常数. 模型变为

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + z_t, \quad z_t = \phi z_{t-1} + a_t, \quad t = 1, \cdots, n \tag{10.6}$$

其中 n 为样本量.

这种情形下，进行吉布斯抽样的一个自然方法是在回归估计和时间序列估计之间进行迭代. 如果已知这个时间序列模型，则我们可以利用最小二乘方法很容易地估计回归模型. 然而，如果回归模型是已知的，则我们可以利用 $z_t = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}$ 得到时间序列 z_t，并利用这个序列估计 AR(1)模型. 因此，我们需要下面的条件后验分布：

$$f(\boldsymbol{\beta} \mid \boldsymbol{Y}, \boldsymbol{X}, \phi, \sigma^2); \quad f(\phi \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\beta}, \sigma^2); \quad f(\sigma^2 \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\beta}, \phi)$$

其中 $\boldsymbol{Y} = (y_1, \cdots, y_n)'$，$\boldsymbol{X}$ 表示解释变量所有观测值的集合.

我们利用共轭先验分布得到条件后验分布的一个显式表达. 先验分布为

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_o, \Sigma_o), \quad \phi \sim N(\phi_o, \sigma_o^2), \quad \frac{v\lambda}{\sigma^2} \sim \chi_v^2 \tag{10.7}$$

这里～又表示分布，$\boldsymbol{\beta}_o$，$\boldsymbol{\Sigma}_o$，λ，v，ϕ_o 和 σ_o^2 是已知量. 在贝叶斯推断中，这些量称为超参数(hyperparameters). 它们的精确值依赖于要处理的问题. 典型地，我们假定$\boldsymbol{\beta}_o=0$，$\phi_o=0$，且 Σ_o 表示具有很大对角元素的对角阵. 假定方程(10.7)中的先验分布相互独立. 这样，我们使用基于参数向量$\boldsymbol{\theta}$的分解的独立先验分布. [406]

条件后验分布 $f(\boldsymbol{\beta} \mid \boldsymbol{Y}, \boldsymbol{X}, \phi, \sigma^2)$可以通过 10.3 节的结果 1a 得到. 具体地，给定 ϕ，定义 $y_{o,t}=y_t-\phi y_{t-1}$，$\boldsymbol{x}_{o,t}=\boldsymbol{x}_t-\phi \boldsymbol{x}_{t-1}$. 利用方程(10.6)，我们有

$$y_{o,t}=\boldsymbol{\beta}'\boldsymbol{x}_{o,t}+a_t, \quad t=2,\cdots,n \tag{10.8}$$

在$\{a_t\}$的假定下，方程(10.8)是一个多元线性回归. 因此，关于参数向量$\boldsymbol{\beta}$的数据信息包在它的最小二乘估计

$$\hat{\boldsymbol{\beta}}=\Big(\sum_{t=2}^{n}\boldsymbol{x}_{o,t}\boldsymbol{x}_{o,t}'\Big)^{-1}\Big(\sum_{t=2}^{n}\boldsymbol{x}_{o,t}y_{o,t}\Big)$$

中，这个最小二乘估计具有多元正态分布

$$\hat{\boldsymbol{\beta}}\sim N\Big[\boldsymbol{\beta},\sigma^2\Big(\sum_{t=2}^{n}\boldsymbol{x}_{o,t}\boldsymbol{x}_{o,t}'\Big)^{-1}\Big].$$

利用结果 1a，给定数据，ϕ 以及 σ^2 的条件下，$\boldsymbol{\beta}$的后验分布是多元正态的. 我们将结果写为

$$(\boldsymbol{\beta} \mid \boldsymbol{Y}, \boldsymbol{X}, \phi, \sigma)\sim N(\boldsymbol{\beta}_*, \Sigma_*) \tag{10.9}$$

其中参数由下式给出

$$\Sigma_*^{-1}=\frac{\sum_{t=2}^{n}\boldsymbol{x}_{o,t}\boldsymbol{x}_{o,t}'}{\sigma^2}+\Sigma_o^{-1}, \quad \boldsymbol{\beta}_*=\Sigma_*\left[\frac{\sum_{t=2}^{n}\boldsymbol{x}_{o,t}\boldsymbol{x}_{o,t}'}{\sigma^2}\hat{\boldsymbol{\beta}}+\Sigma_o^{-1}\boldsymbol{\beta}_o\right].$$

下面考虑给定$\boldsymbol{\beta}$，σ^2 和数据条件下，ϕ 的条件后验分布. 因为$\boldsymbol{\beta}$给定，所以对所有的 t，可以计算 $z_t=y_t-\boldsymbol{\beta}'\boldsymbol{x}_t$，并考虑 AR(1)模型 $z_t=\phi z_{t-1}+a_t$，$t=2$，…，n. 有关 ϕ 的似然函数信息包含在最小二乘估计

$$\hat{\phi}=\Big(\sum_{t=2}^{n}z_{t-1}^2\Big)^{-1}\Big(\sum_{t=2}^{n}z_{t-1}z_t\Big)$$

[407]

中，它服从均值为 ϕ、方差 $\sigma^2\Big(\sum_{t=2}^{n}z_{t-1}^2\Big)^{-1}$ 的正态分布. 根据结果 1，ϕ 的后验分布也是正态，均值 ϕ_*，方差 σ_*^2，其中

$$\sigma_*^{-2}=\frac{\sum_{t=2}^{n}z_{t-1}^2}{\sigma^2}+\sigma_o^{-2}, \quad \phi_*=\sigma_*^2\left[\frac{\sum_{t=2}^{n}z_{t-1}^2}{\sigma^2}\hat{\phi}+\sigma_o^{-2}\phi_o\right] \tag{10.10}$$

最后，转到给定 β，ϕ 和数据条件下，σ^2 的后验分布. 因为 β 和 ϕ 已知，可以计算

$$a_t=z_t-\phi z_{t-1}, \quad z_t=y_t-\beta' x_t, \quad t=2,\cdots,n$$

由 10.3 节的结果 8，σ^2 的后验分布是逆 χ^2 分布，即

$$\frac{v\lambda+\sum_{t=2}^{n}a_t^2}{\sigma^2}\sim\chi^2_{v+(n-1)} \tag{10.11}$$

其中 χ_k^2 表示自由度为 k 的 χ^2 分布.

利用方程(10.9)～(10.11)的三个条件后验分布，我们可以通过吉布斯抽样估计方程(10.6)，步骤如下：

1)在方程(10.7)中指定先验分布的超参数值.

2)指定$\boldsymbol{\beta}$，ϕ 和 σ^2 的任意初始值(如不带时间序列误差的 β 的普通最小二乘估计).

3)利用方程(10.9)的多元正态分布抽取$\boldsymbol{\beta}$的一个随机实现.

4)利用方程(10.10)的多元正态分布抽取 ϕ 的一个随机实现.

5)利用方程(10.11)的多元 χ^2 分布抽取 σ^2 的一个随机实现.

重复步骤 3～5 许多次迭代；得到吉布斯样本. 然后利用样本均值作为模型(10.6)中参数的点估计.

例 10.1 重温第 2 章中关于美国周利率的例子. 数据是从 1962 年 1 月 5 日到 1999 年 9 月 10 日的固定期限为 1 年和 3 年国库券利率，来自于 St Louis 的联邦储备银行. 因为单位根非平稳性，独立和不独立的变量为

1)$c_{3t}=r_{3t}-r_{3,t-1}$，3 年期利率的周变化；

408

2)$c_{1t}=r_{1t}-r_{1,t-1}$，1 年期利率的周变化，

这里原始的利率 r_{it}是以百分比测量的. 第 2 章中，我们对数据采用了一个带 MA(1)误差的线性回归模型. 这里我们考虑误差过程为 AR(2)模型. 利用传统方法，我们得到模型

$$c_{3t} = 0.0002 + 0.782c_{1t} + z_t \quad z_t = 0.205z_{t-1} - 0.068z_{t-2} + a_t \qquad (10.12)$$

其中$\hat{\sigma}_a=0.067$. 方程(10.12)中系数估计的标准误差分别为 0.001 7，0.008，0.023 和 0.023. 残差 ACF 除了在延迟 6 处的边际显著外，前面的模型看上去是充分的.

将模型写为：

$$c_{3t} = \beta_0 + \beta_1 c_{1t} + z_t, \quad z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t \qquad (10.13)$$

其中$\{a_t\}$是独立的 $N(0, \sigma^2)$随机变量的序列，我们通过吉布斯抽样估计参数. 前面使用的先验分布为

$$\boldsymbol{\beta} \sim N(\mathbf{0}, 4\boldsymbol{I}_2), \quad \phi \sim N[\mathbf{0}, \text{diag}(0.25, 0.16)], \quad (v\lambda)/\sigma^2 = (10 \times 0.1)/\sigma^2 \sim \chi_{10}^2$$

这里 $\boldsymbol{I}_2$ 是 2×2 单位矩阵. 初始参数估计是通过 OLS 方法得到的(即利用一个 2 步程序，先拟合一个线性回归模型，然后对回归残差拟合一个 AR(2)模型). 因为样本量 1 966 很大，所以初始估计接近于方程(10.12)给出的结果. 我们重复吉布斯抽样迭代 2 100 次，但抛弃前 100 个迭代结果. 表 10-1 给出了参数的后验均值和标准差. 图 10-1 显示了每一个参数的边际后验分布的直方图.

表 10-1

参数	β_0	β_1	ϕ_1	ϕ_2	σ
均值	0.025	0.784	0.305	0.032	0.074
标准差	0.024	0.009	0.089	0.087	0.003

注：用 2 100 次迭代的吉布斯抽样来估计的模型(10.13)的后验均值和标准差. 结果基于后 2 000 个迭代，先验分布在前文中给出.

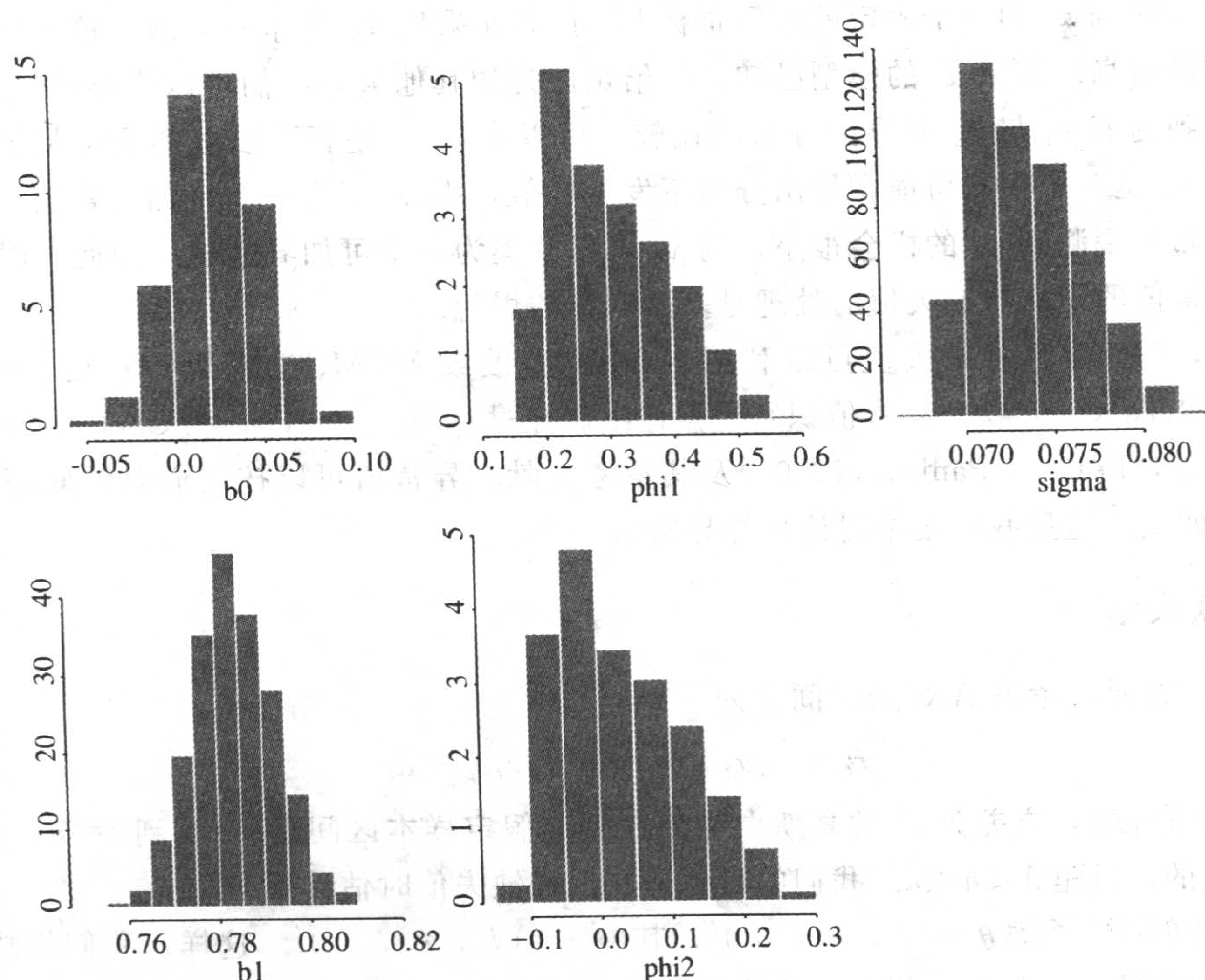


图 10-1 模型(10.13)的 2 100 次迭代的吉布斯抽样直方图. 结果基于后 2 000 个迭代，先验分布及初始参数值在前文中给出

我们以不同的初始值重复吉布斯抽样，都得到相似结果. 吉布斯抽样看上去是收敛的. 由表 10-1，除了 z_{t-2} 的系数外，后验均值都接近于方程(10.12)的估计. 然而，ϕ_1 和 ϕ_2 的后验标准误差相对较大，显示了这两个估计的不确定性. 图 10-1 的直方图是有用的. 特别地，它们显示了$\hat{\phi}_1$ 和$\hat{\phi}_2$的分布并不收敛到渐近正态；分布是右偏的. 然而，$\hat{\beta}_0$ 和$\hat{\beta}_1$ 的渐近正态性看似是合理的.

10.6 缺失值和异常值

本节中，我们讨论处理缺失值和识别异常值的 MCMC 方法. 令$\{y_t\}_{t=1}^{n}$表示观测的时间序列. 数据点 y_h 是一个可加的异常值，如果

$$y_t = \begin{cases} x_h + \omega & 若\ t = h \\ x_t & 其他 \end{cases} \tag{10.14}$$

这里 ω 是异常值的大小，x_t 是没有异常值的时间序列. 可加异常值的例子包括记录误差(如打字误差和测量误差). 异常值可以严重影响时间序列分析，因为它们可以导致参数估计中的较大偏差，并导致模型的错误指定.

409 ~ 410

考虑时间序列 x_t 和一个固定的时间指标 h. 我们如果把 x_h 当作一个缺失值，则可以了解很多关于它的信息. 如果 x_t 的模型已知，则给定序列的其他值，我们可以推导出 x_h 的条件分布. 把观测到的值 y_h 与 x_h 的导出分布相比较，可以决定 y_h 是否可以分类为可加异常值. 具体地，如果 y_h 是一个很有可能在导出分布下发生的值，则 y_h 不是一个可加异常值. 然而，如果在导出分布下观测到 y_h 的机会很小，则 y_h 可以分类为一个可加异常值. 因此，时间序列分析中可加异常值的识别与缺失值的处理是基于同样的思想.

文献中，时间序列的缺失值可以利用 Kalman 滤波或 MCMC 方法处理；见 Jones(1980) McCulloch 和 Tsay(1994). 异常值识别也被仔细研究过；Chang，Tiao 和 Chen (1988)，Tsay (1988)，Tsay，Peña 和 Pankratz(2000)及其参考文献. 异常值可以按它们对时间序列的影响的性质分为四类. 这里我们集中讨论可加异常值.

10.6.1 缺失值

为了表示方便，考虑 AR(p)时间序列

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t \tag{10.15}$$

其中$\{a_t\}$是均值为 0，方差为 σ^2 的高斯白噪声序列. 假定样本区间是 $t=1$ 到 $t=n$，但是观测值 x_h 是缺失的，这里 $1<h<n$. 我们的目标是在出现缺失值时估计模型.

在这个特例中，参数$\boldsymbol{\theta}=(\boldsymbol{\phi}', x_h, \sigma^2)'$，其中 $\boldsymbol{\phi}=(\phi_1, \cdots, \phi_p)'$. 这样，我们将缺失值 x_h 看作一个未知参数. 如果我们假定先验分布为

$$\boldsymbol{\phi} \sim N(\boldsymbol{\phi}_o, \boldsymbol{\Sigma}_o), \quad x_h \sim N(\mu_o, \sigma_o^2), \quad \frac{v\lambda}{\sigma^2} \sim \chi_v^2$$

其中超参数是已知的，则条件后验分布 $f(\boldsymbol{\phi} \mid \boldsymbol{X}, x_h, \sigma^2)$和 $f(\sigma^2 \mid \boldsymbol{X}, x_h, \boldsymbol{\phi})$正好是前一节中给出的那些，其中 $\boldsymbol{X}$ 表示观测到的数据. 条件后验分布 $f(x_h \mid \boldsymbol{X}, \boldsymbol{\phi}, \sigma^2)$是均值为 μ_* 和方差 σ_h^2 的 1 元正态分布. 这两个参数可以利用线性回归模型得到. 具体地说，给定模型和数据，x_h 仅仅与$\{x_{h-p}, \cdots, x_{h-1}, x_{h+1}, \cdots, x_{h+p}\}$相关. 记住 x_h 是一个未知参数，我们可以将这个关系如下写出：

1)对 $t=h$，模型说明 $x_h=\phi_1 x_{h-1}+\cdots+\phi_p x_{h-p}+a_h$，

令 $y_h=\phi_1 x_{h-1}+\cdots+\phi_p x_{h-p}$，$b_h=-a_h$，则前面的方程可以写成 $y_h=x_h+b_h=\phi_0 x_h+b_h$ 其中 $\phi_0=1$.

411

2)对 $t=h+1$，我们有

$$x_{h+1} = \phi_1 x_h + \phi_2 x_{h-1} + \cdots + \phi_p x_{h+1-p} + a_{h+1},$$

令 $y_{h+1}=x_{h+1}-\phi_2 x_{h-1}-\cdots-\phi_p x_{h+1-p}$且 $b_{h+1}=a_{h+1}$则前面的方程写为 $y_{h+1}=\phi_1 x_h+b_{h+1}$.

3)一般地，对 $t=h+j$，$j=1, \cdots, p$，我们有

$$x_{h+j} = \phi_1 x_{h+j-1} + \cdots + \phi_j x_h + \phi_{j+1} x_{h-1} + \cdots + \phi_p x_{h+j-p} + a_{h+j}$$

令 $y_{h+j}=x_{h+j}-\phi_1 x_{h+j-1}-\cdots-\phi_{j-1} x_{h+1}-\phi_{j+1} x_{h-1}-\cdots-\phi_p x_{h+j-p}$且 $b_{h+j}=a_{h+j}$. 前面方程简化为 $y_{h+j}=\phi_j x_h+b_{h+j}$.

因此，对一个 AR(p)模型，缺失值 x_h 与模型有关，并且在 $p+1$ 个方程中的数据为

$$y_{h+j} = \phi_j x_h + b_{h+j}, \qquad j = 0, \cdots, p, \tag{10.16}$$

其中 $\phi_0 = 1$. 因为正态分布关于它的均值对称，所以 a_h 和 a_{-h} 具有同样的分布. 因此，方程(10.16)是一个具有 $p+1$ 个数据点的特殊的简单线性回归模型. x_h 的最小二乘估计及其方差为

$$\hat{x}_h = \frac{\sum_{j=0}^{p} \phi_j y_{h+j}}{\sum_{j=0}^{p} \phi_j^2}, \quad \text{Var}(\hat{x}_h) = \frac{\sigma^2}{\sum_{j=0}^{p} \phi_j^2}.$$

例如，当 $p=1$ 时，我们有 $\hat{x}_h = \frac{\phi_1}{1+\phi_1^2}(x_{h-1} + x_{h+1})$，称为 x_h 的过滤值. 因为高斯 AR(1)模型是时间可逆的，所以对 x_h 的两个相邻观测运用等权重得到过滤值.

最后，利用 10.3 节的结果 1，我们得到 x_h 的后验分布是均值为 μ_* 方差为 σ_*^2 的正态分布，其中

$$\mu_* = \frac{\sigma^2 \mu_o + \sigma_o^2 \left(\sum_{j=0}^{p} \phi_j^2 \right) \hat{x}_h}{\sigma^2 + \sigma_o^2 \left(\sum_{j=0}^{p} \phi_j^2 \right)}, \quad \sigma_*^2 = \frac{\sigma^2 \sigma_o^2}{\sigma^2 + \sigma_o^2 \sum_{j=0}^{p} \phi_j^2} \tag{10.17}$$

缺失值可能成堆发生，导致了多个连续的缺失值的情形. 这些缺失值可以用两种方法处理. 第一种，我们可以直接推广前面的方法得到多个过滤值的解. 例如，考虑 x_h 和 x_{h+1} 缺失的情形. 这些缺失值与 $\{x_{h-p}, \cdots, x_{h-1}; x_{h+2}, \cdots, x_{h+p+1}\}$ 有关. 我们可以用与以前同样的方式定义一个响应变量 y_{h+j}，从而建立一个参数为 x_h 和 x_{h+1} 的多元线性回归. 然后利用最小二乘法得到 x_h 和 x_{h+1} 的估计. 与指定的先验分布相结合，我们有对 $(x_h, x_{h+1})'$ 的一个二元正态后验分布. 在吉布斯抽样中，这种方法可联合地抽取相连的缺失值. 第二，我们可以在吉布斯迭代中多次利用方程(10.17)中单个缺失值的结果. 再次考虑缺失 x_h 和 x_{h+1} 的情形. 我们可以分别地利用后验分布 $f(x_h \mid \boldsymbol{X}, x_{h+1}, \boldsymbol{\phi}, \sigma^2)$ 和 $f(x_{h+1} \mid \boldsymbol{X}, x_h, \boldsymbol{\phi}, \sigma^2)$. 在吉布斯抽样中，这意味着我们一次抽取一个缺失值. [412]

因为 x_h 和 x_{h+1} 在时间序列中是相关的，所以，在吉布斯抽样中更愿意联合地抽取它们，当连续缺失值的数量很大时尤其如此. 如果缺失值的数量很小，则一次抽取一个缺失值进行得很好.

注释：*在前面讨论中，我们假定 $h-p \geqslant 1$ 且 $h+p \leqslant n$. 如果 h 接近于样本时间区间的终点，则在线性回归模型中可以利用的数据点的数量必须调整.*

10.6.2 异常值的识别

方程(10.14)中可加异常值的识别在 MCMC 框架中变得很直接. 除了具有相似大小的一堆可加异常值的情形，McCulloch 和 Tsay(1994)的简单吉布斯抽样运行都很好；见 Justel，Peña 和 Tsay(2001). 我们再次利用 AR 模型来解释这个问题. 这个方法当利用 Metropolis-Hasting 算法或格子吉布斯抽取非线性参数时，与其他的时间序列模型应用得同样好.

假定观测的时间序列为 y_t，它可能包含位置和大小都未知的一些可加异常值. 我们将 y_t 的模型写为

$$y_t = \delta_t\beta_t + x_t, \qquad t = 1,\cdots,n \tag{10.18}$$

其中$\{\delta_t\}$是独立的伯努利随机变量序列，满足 $P(\delta_t=1)=\varepsilon$，$P(\delta_t=0)=1-\varepsilon$，$\varepsilon$ 是 0 到 1 之间的常数，$\{\beta_t\}$是来自给定分布中的一个独立随机变量序列，x_t 是无异常值的 AR(p)时间序列

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t$$

其中$\{a_t\}$是均值为 0、方差为 σ^2 的高斯白噪声. 虽然这个模型看起来复杂，但是它允许可加异常值在每个时间点发生. 每一个观测是异常值的机会为 ε.

413

在方程(10.18)的模型下，我们有 n 个数据点，但是有 $2n+p+3$ 个参数，即 $\boldsymbol{\phi}=(\phi_0, \cdots, \phi_p)'$，$\boldsymbol{\delta}=(\delta_1, \cdots, \delta_n)'$，$\boldsymbol{\beta}=(\beta_1, \cdots, \beta_n)'$，$\sigma^2$ 和 ε. 二值参数 δ_t 由 ε 控制，$\beta_t s$ 由指定的分布确定. 参数$\boldsymbol{\delta}$与$\boldsymbol{\beta}$是利用数据扩张的思想引进的，δ_t 表示在时刻 t 异常值出现或不出现，β_t 表示当时刻 t 异常值出现时它的大小.

假定先验分布为 $\boldsymbol{\phi}\sim N(\boldsymbol{\phi}_o, \Sigma_o)$，$\frac{v\lambda}{\sigma^2}\sim\chi^2_v$，$\varepsilon\sim \text{beta}(\gamma_1, \gamma_2)$，$\beta_t\sim N(0, \xi^2)$，其中超参数是已知的. 这些是共轭先验分布. 为了对具有异常值识别的模型估计实施吉布斯抽样，我们需要考虑条件后验分布：$f(\boldsymbol{\phi}\mid \boldsymbol{Y}, \boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2)$，$f(\delta_h \mid \boldsymbol{Y}, \boldsymbol{\delta}_{-h}, \boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2)$，$f(\beta_h \mid \boldsymbol{Y}, \boldsymbol{\delta}, \boldsymbol{\beta}_{-h}, \boldsymbol{\phi}, \sigma^2)$，$f(\varepsilon \mid \boldsymbol{Y}, \boldsymbol{\delta})$，$f(\sigma^2 \mid \boldsymbol{Y}, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\beta})$，这里 $1\leqslant h\leqslant n$，$\boldsymbol{Y}$ 表示数据，$\boldsymbol{\theta}_{-i}$表示剔除$\boldsymbol{\theta}$的第 i 个元素后的向量.

以$\boldsymbol{\delta}$和$\boldsymbol{\beta}$为条件，没有异常值的时间序列可以通过 $x_t=y_t-\delta_t\beta_t$ 得到，数据中关于 $\boldsymbol{\phi}$ 的信息包含在它的最小二乘估计

$$\hat{\boldsymbol{\phi}} = \left(\sum_{t=p+1}^{n} \boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right)^{-1}\left(\sum_{t=p+1}^{n} \boldsymbol{x}_{t-1}x_t\right)$$

中，其中 $\boldsymbol{x}_{t-1}=(1, x_{t-1}, \cdots, x_{t-p})'$，它服从正态分布，均值为 $\boldsymbol{\phi}$，协方差矩阵为

$$\hat{\Sigma}=\sigma^2\left(\sum_{t=p+1}^{n} \boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\right)^{-1}.$$

因此，$\boldsymbol{\phi}$ 的条件后验分布是多元正态的，均值为 $\boldsymbol{\phi}_*$ 和协方差矩阵Σ_*由方程(10.9)给出，其中用 $\boldsymbol{\phi}$ 代替$\boldsymbol{\beta}$，$\boldsymbol{x}_{t-1}$代替 $\boldsymbol{x}_{o,t}$. 类似地，σ^2 的条件后验分布是逆 χ^2 分布，即

$$\frac{v\lambda + \sum_{t=p+1}^{n} a_t^2}{\sigma^2} \sim \chi^2_{v+(n-p)},$$

其中 $a_t=x_t-\boldsymbol{\phi}'\boldsymbol{x}_{t-1}$，$x_t=y_t-\delta_t\beta_t$.

414

δ_h 的条件后验分布可以如下得到. 第一，δ_h 仅仅与$\{y_j, \beta_j\}_{j=h-p}^{h+p}$相关，$\{\delta_j\}_{j=h-p}^{h+p}$与 $j\neq h$，$\boldsymbol{\phi}$ 和 σ^2 相关. 更具体地，我们有 $x_j=y_j-\delta_j\beta_j$，$j\neq h$；第二，可以假定 x_h 有两个可能的值：如果 $\delta_h=1$，则 $x_h=y_h-\beta_h$；否则，$x_h=y_h$. 定义 $w_j=x_j^*-\phi_0-\phi_1x_{j-1}^*-\cdots-\phi_px_{j-p}^*(j=h, \cdots, h+p)$，其中，如果 $j\neq h$，$x_j^*=x_j$ 且 $x_h^*=y_h$. x_h 的两个可能取值给出了两种情形：

- 情形Ⅰ：$\delta_h=0$. 这里第 h 个观测不是一个异常值且 $x_h^*=y_h=x_h$，从而对 $j=h, \cdots, h+p$，$w_j=a_j$. 换句话说,我们有 $w_j\sim N(0, \sigma^2)$，$j=h, \cdots, h+p$.

- 情形Ⅱ：$\delta_h=1$. 现在第 h 个观测是一个异常值，且 $x_h^*=y_h=x_h+\beta_h$，则前面定义的 w_j 被 β_h 污损了. 事实上，我们有 $w_h\sim N(\beta_h,\sigma^2)$，$w_j\sim N(-\phi_{j-h}\beta_h,\sigma^2)$，$j=h+1,\cdots,h+p$. 如果我们定义 $\psi_0=-1$，$\psi_i=\phi_i$ 对 $i=1,\cdots,p$，则对 $j=h,\cdots,h+p$，我们有 $w_j\sim N(-\psi_{j-h}\beta_h,\sigma^2)$.

根据前面的讨论，我们可以概括如下：

1)情形Ⅰ：以概率 $1-\varepsilon$，$\delta_h=0$. 这种情形下，对 $j=h,\cdots,h+p$，$w_j\sim N(0,\sigma^2)$；

2)情形Ⅱ：以概率 ε，$\delta_h=1$. 这里对 $j=h,\cdots,h+p$，$w_j\sim N(-\psi_{j-h}\beta_h,\sigma^2)$.

因为有 n 个数据点，所以 j 不能大于 n. 令 $m=\min(n,h+p)$，从而 δ_h 的后验分布为

$$P(\delta_h=1\mid \boldsymbol{Y},\boldsymbol{\delta}_{-h},\boldsymbol{\beta},\boldsymbol{\phi},\sigma^2)$$

$$=\frac{\varepsilon\exp\left[-\sum_{j=h}^{m}(w_j+\psi_{j-h}\beta_h)^2/(2\sigma^2)\right]}{\varepsilon\exp\left[-\sum_{j=h}^{m}(w_j+\psi_{j-h}\beta_h)^2/(2\sigma^2)\right]+(1-\varepsilon)\exp\left[-\sum_{j=h}^{m}w_j^2/(2\sigma^2)\right]} \quad (10.19)$$

这个后验分布仅仅是比较两种情形下似然函数的权重，而权重是每种情况的概率. 最后，β_h 的后验分布如下： 415

- 如果 $\delta_h=0$，则 y_h 不是一个异常值，且 $\beta_h\sim N(0,\xi^2)$.
- 如果 $\delta_h=1$，则 y_h 由大小为 β_h 的异常值污染. 对 $j=h,h+1,\cdots,\min(h+p,n)$，前面定义的变量 w_j 包含了 β_h 的信息，特别是我们有 $w_j\sim N(-\psi_{j-h}\beta_h,\sigma^2)$. 这个信息可以加入一个线性回归的框架 $w_j=-\psi_{j-h}\beta_h+a_j$，$j=h,h+1,\cdots,\min(h+p,n)$. 因此，这个信息包含在最小二乘估计

$$\hat{\beta}_h=\frac{\sum_{j=h}^{m}-\psi_{j-h}w_j}{\sum_{j=h}^{m}\psi_{j-h}^2},\quad m=\min(h+p,n),$$

它是均值为 β_h、方差 $\sigma^2/\left(\sum_{j=h}^{m}\psi_{j-h}^2\right)$ 的正态分布. 由结果1，β_h 的后验分布是均值为 β_h^*，方差 σ_{h*}^2 的正态分布，其中

$$\beta_h^*=\frac{-\left(\sum_{j=h}^{m}\psi_{j-h}w_j\right)\xi^2}{\sigma^2+\left(\sum_{j=h}^{m}\psi_{j-h}^2\right)\xi^2},\quad \sigma_{h*}^2=\frac{\sigma^2\xi^2}{\sigma^2+\left(\sum_{j=h}^{m}\psi_{j-h}^2\right)\xi^2}.$$

例 10.2 考虑从 1988 年 3 月 18 日至 1999 年 9 月 10 日美国的具有固定期限的 3 年期国库券利率的周变化序列，共有 600 个观测值. 利率以百分比表示，是例 10.1 中的响应变量 c_{3t} 的子序列. 图 10-2a 显示了该时间序列. 如果对序列利用 AR 模型，则其 PACF 建议一个 AR(3)模型，且我们得到

$$c_{3t}=0.227c_{3,t-1}+0.006c_{3,t-2}+0.114c_{3,t-2}+a_t,\hat{\sigma}_a^2=0.0128,$$

其中系数的标准差分别为 0.041，0.042 和 0.041. 残差的 L-B 统计量为 $Q(12)=11.4$，它在 5%水平下不显著.

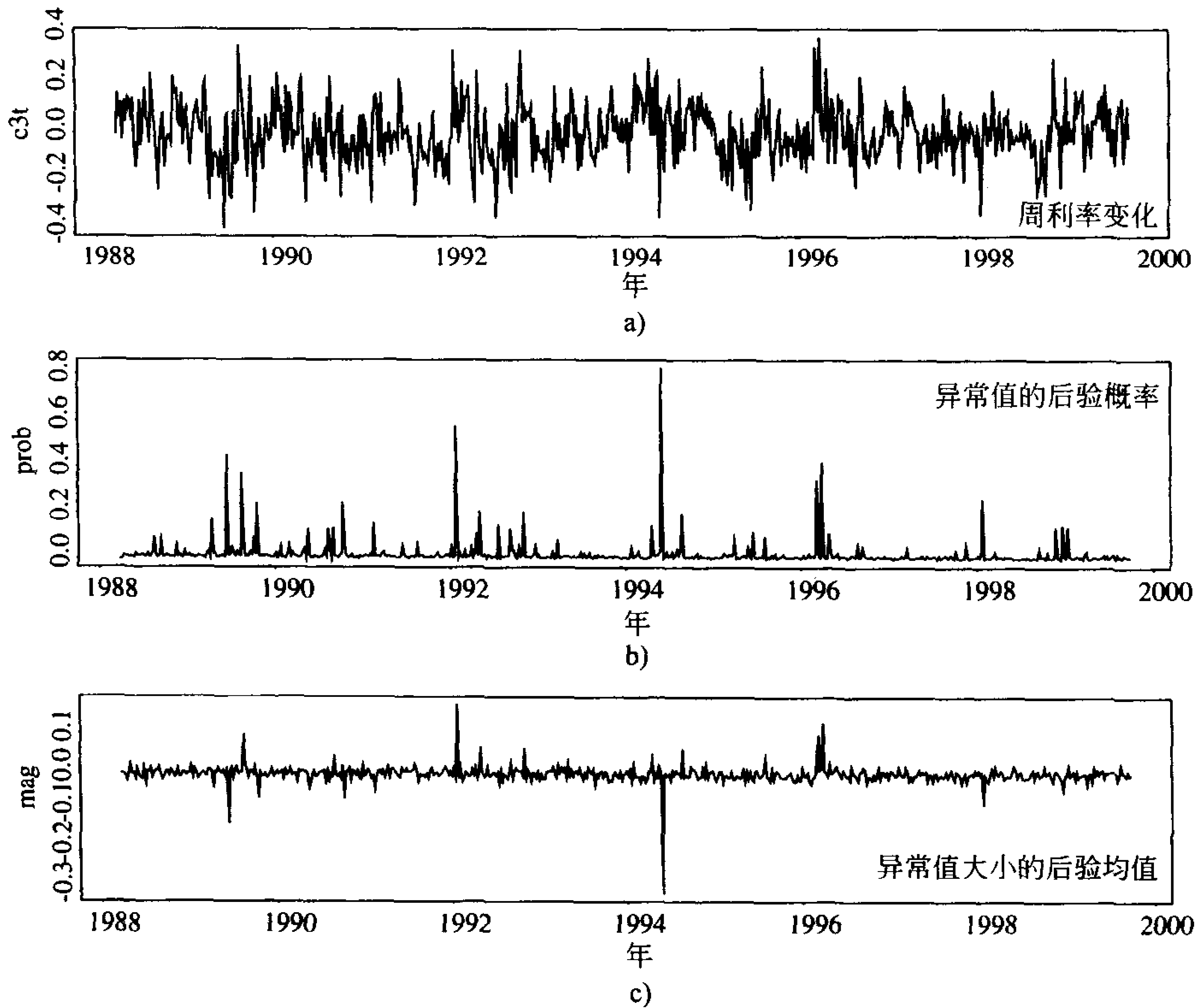


图 10-2 从 1988 年 3 月 18 日至 1999 年 9 月 10 日美国的具有固定期限的 3 年期国库券利率的周变化序列时间图. a)数据，b)为异常值的后验概率，c)异常值大小的后验均值. 估计基于 1 050 个迭代的吉布斯抽样，但前 50 个迭代被删去

下一步利用吉布斯抽样来估计这个 AR(3)模型，并同时识别可能的可加异常值. 所用的先验分布为

$$\phi\sim N(\mathbf{0},0.25\boldsymbol{I}_3),\quad \frac{v\lambda}{\sigma^2}=\frac{5\times0.00256}{\sigma^2}\sim\chi_5^2,\quad \gamma_1=5,\quad \gamma_2=95,\quad \xi^2=0.1$$

其中 $0.00256\approx\hat{\sigma}^2/5$ 且 $\xi^2\approx9\hat{\sigma}^2$. 可加异常值的期望数量为 5%. 利用初始值 $\varepsilon=0.05$，$\sigma^2=0.012$，$\phi_1=0.2$，$\phi_2=0.02$，$\phi_3=0.1$，我们运行吉布斯抽样 1 050 个迭代，但是将前 50 个迭代的结果去掉. 利用系数的后验均值作为参数估计，我们得到拟合的模型

$$c_{3t}=0.252c_{3,t-1}+0.003c_{3,t-2}+0.110c_{3,t-2}+a_t,\quad \hat{\sigma}^2=0.0118,$$

其中参数的后验标准差分别为 0.046，0.045，0.046 和 0.000 8. 这样，吉布斯抽样产生的结

果类似于最大似然方法的结果. 图 10-2b 显示了可加异常值的每一个观测的后验概率的时间图，图 10-2c 画出了异常值大小的后验均值图形. 由概率图，有些观测值取异常值的概率很高. 特别地，$t=323$ 的概率为 0.83，相应的异常值大小的后验均值 −0.304. 当 c_{3t} 从 0.24 变化到 −0.34(即大约是两星期内周利率降低 0.6%)时，这个点对应于 1994 年 5 月 20 日. 异常值的第二个最高后验概率的点为 $t=201$，是 1992 年 1 月 17 日. 列出的后验概率为 0.58，且估计的异常值大小为 0.176. 在第二个点，从 −0.02 变化到 0.33，相应于周利率的一个大约 0.35%的跳跃.

416～417

注释： 通过吉布斯抽样的异常值识别要求高强度的计算，但这个方法对模型参数和异常值实施了一个联合估计. 而传统的异常值辨识方法是将估计从辨识中分离. 它在计算上更快，但是，当多个异常值出现时，可能产生错误的辨识. 对于例 10.2 种的数据，SCA 程序也识别了 $t=323$ 和 $t=201$，作为两个最显著的可加异常值. 所估的异常值大小分别为 −0.39 和 0.36.

10.7 随机波动率模型

MCMC 方法的一个重要的金融应用是随机波动率模型的估计，见 Jacquier，Polson 和 Rossi(1994)及参考文献. 我们以一个一元随机波动率模型开始. 资产收益率的均值和波动率方程为

$$r_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_p x_{pt} + a_t, \quad a_t = \sqrt{h_t}\epsilon_t \tag{10.20}$$

$$\ln h_t = \alpha_0 + \alpha_1 \ln h_{t-1} + v_t, \tag{10.21}$$

其中$\{x_{it} \mid i=1, \cdots, p\}$是时刻 $t-1$ 可以得到的解释变量，β_j 是参数，$\{\epsilon_t\}$是均值为 0、方差为 1 的高斯白噪声序列，$\{v_t\}$也是高斯白噪声序列，其均值为 0、方差为 σ_v^2，且$\{\epsilon_t\}$和 $\{v_t\}$是独立的. 用对数变换保证对所有 t，h_t 是正的. 解释变量 x_{it} 可能包含收益率的延迟值(如 $x_{it}=r_{t-i}$). 方程(10.21)中，我们假定 $|\alpha_1|<1$，使得对数波动率过程 $\ln h_t$ 平稳. 如果必要，可以对 $\ln h_t$ 利用更高阶 AR(p)模型.

记均值方程的系数向量为$\boldsymbol{\beta}=(\beta_0, \beta_1, \cdots, \beta_p)'$，记波动率方程的参数向量为$\boldsymbol{\omega}=(\alpha_0, \alpha_1, \sigma_v^2)'$，假如$\boldsymbol{R}=(r_1, \cdots, r_n)'$是观测到的收益率集合，而 $\boldsymbol{X}$ 是解释变量的集合. 令 $\boldsymbol{H}=(h_1, \cdots, h_n)'$为无法观测到的波动率向量. 这里$\boldsymbol{\beta}$和$\boldsymbol{\omega}$ 是模型的“传统”参数，而 $\boldsymbol{H}$ 是一个辅助变元. 如果通过最大似然估计法对模型进行估计，则会很复杂，因为似然函数是以下 n 维 $\boldsymbol{H}$ 分布的混合：

$$f(\boldsymbol{R} \mid \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\omega}) = \int f(\boldsymbol{R} \mid \boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{H}) f(\boldsymbol{H} \mid \boldsymbol{\omega}) \mathrm{d}\boldsymbol{H}.$$

然而，在贝叶斯框架下，波动率向量 $\boldsymbol{H}$ 由扩充参数组成. 在 $\boldsymbol{H}$ 条件下，我们可以关注概率分布函数 $f(\boldsymbol{R} \mid \boldsymbol{H}, \boldsymbol{\beta})$和 $f(\boldsymbol{H} \mid \boldsymbol{\omega})$还有先验分布 $p(\boldsymbol{\beta}, \boldsymbol{\omega})$. 假设先验分布可以分解成 $p(\boldsymbol{\beta}, \boldsymbol{\omega})=$

418

$p(\boldsymbol{\beta})p(\boldsymbol{\omega})$，即均值和波动率方程的先验分布是独立的. 估计方程(10.20)和方程(10.21)中的随机波动率的一个吉布斯抽样方法涉及到从以下条件后验分布中抽取随机样本：

$$f(\boldsymbol{\beta} \mid \boldsymbol{R},\boldsymbol{X},\boldsymbol{H},\boldsymbol{\omega}),\quad f(\boldsymbol{H} \mid \boldsymbol{R},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{\omega}),\quad f(\boldsymbol{\omega} \mid \boldsymbol{R},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{H})$$

下面，我们给出吉布斯抽样器在实际应用中的操作细节.

10.7.1 一元模型的估计

给定 $\boldsymbol{H}$，式(10.20)中的均值方程是一个非齐次的线性回归方程. 方程两边同时除以$\sqrt{h_t}$，我们可以将模型改写为：

$$r_{o,t} = \boldsymbol{x}'_{o,t}\boldsymbol{\beta} + \epsilon_t,\qquad t = 1,\cdots,n \tag{10.22}$$

其中 $r_{o,t}=r_t/\sqrt{h_t}$和 $x_{o,t}=x_t/\sqrt{h_t}$，且 $x_t=(1,\ x_{1t},\ \cdots,\ x_{pt})'$是解释变量的向量. 假如$\boldsymbol{\beta}$的先验分布是均值为$\boldsymbol{\beta}_o$，协方差阵为 $\boldsymbol{A}_o$ 的多元正态分布，则$\boldsymbol{\beta}$的后验分布也是多元正态分布，均值是$\boldsymbol{\beta}_*$，协方差阵是 $\boldsymbol{A}_*$. 这两个量可以像以前一样通过结果 1a 得到，他们是：

$$A_*^{-1} = \sum_{t=1}^{n}\boldsymbol{x}_{o,t}\boldsymbol{x}'_{o,t}+\boldsymbol{A}_o^{-1},\quad \boldsymbol{\beta}_* = \boldsymbol{A}_*\left(\sum_{t=1}^{n}\boldsymbol{x}_{o,t}r_{o,t}+\boldsymbol{A}_o^{-1}\boldsymbol{\beta}_o\right)$$

其中和式理解为从 $p+1$ 开始求和，如果 r_{t-p}是在解释变量中使用的最大延迟收益率.

波动率向量 $\boldsymbol{H}$ 是逐个元素地抽样的. 必需的条件后验分布是 $f(h_t \mid \boldsymbol{R},\ \boldsymbol{X},\ \boldsymbol{H}_{-t},\ \boldsymbol{\beta},\ \boldsymbol{\omega})$，它是由 a_t 的正态分布和波动率的对数正态分布产生的，

$$\begin{aligned} &f(h_t \mid \boldsymbol{R},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{H}_{-t},\boldsymbol{\omega})\\ &\propto f(a_t \mid h_t,r_t,\boldsymbol{x}_t,\boldsymbol{\beta})f(h_t \mid h_{t-1},\boldsymbol{\omega})f(h_{t+1} \mid h_t,\boldsymbol{\omega})\\ &\propto h_t^{-0.5}\exp[-(r_t-\boldsymbol{x}'_t\boldsymbol{\beta})^2/(2h_t)]h_t^{-1}\exp[-(\ln h_t-\mu_t)^2/(2\sigma^2)]\\ &\propto h_t^{-1.5}\exp[-(r_t-\boldsymbol{x}'_t\boldsymbol{\beta})^2/(2h_t)-(\ln h_t-\mu_t)^2/(2\sigma^2)] \end{aligned} \tag{10.23}$$

其中 $\mu_t=[\alpha_0(1-\alpha_1)+\alpha_1(\ln h_{t+1}+\ln h_{t-1})]/(1+\alpha_1^2)$和 $\sigma^2=\sigma_v^2/(1+\alpha_1^2)$. 这里我们使用了以下性质：(a)$a_t \mid h_t \sim N(0,\ h_t)$；(b)$\ln h_t \mid \ln h_{t-1} \sim N(\alpha_0+\alpha_1\ln h_{t-1},\ \sigma_v^2)$；(c)$\ln h_{t+1} \mid \ln h_t \sim N(\alpha_0+\alpha_1\ln h_t,\ \sigma_v^2)$；(d)$d\ln h_t=h_t^{-1}dh_t$，其中 d 表示微分算子和(e)等式

419

$$(x-a)^2A+(x-b)^2C=(x-c)^2(A+C)+(a-b)^2AC/(A+C)$$

其中 $c=(Aa+Cb)/(A+C)$，如果 $A+C\neq0$. 这个等式是 Box 和 Tiao(1973，第 418 页)中的引理 1 的一个标量版本. 在我们的应用中，$A=1$，$a=\alpha_0+\ln h_{t-1}$，$C=\alpha_1^2$ 和 $b=(\ln h_{t+1}-\alpha_0)/\alpha_1$. $(a-b)^2AC/(A+C)$项不包含随机变量 h_t，从而在条件后验分布的导数的积分中被消掉了. Jacquier，Polson 和 Rossi(1994)用 Metropolis 算法来抽取 h_t 的样本. 这一部分我们用格子吉布斯抽样器，而 h_t的取值范围是 r_t 的无条件样本方差的倍数.

为了抽取$\boldsymbol{\omega}$ 的随机样本，我们将参数分解为$\boldsymbol{\alpha}=(\alpha_0,\ \alpha_1)'$和 σ_v^2. $\boldsymbol{\omega}$ 的先验分布也可相应地分解(即，$p(\boldsymbol{\omega})=p(\boldsymbol{\alpha})p(\sigma_v^2)$). 我们需要的条件后验分布是：

- $f(\boldsymbol{\alpha} \mid \boldsymbol{Y},\ \boldsymbol{X},\ \boldsymbol{H},\ \boldsymbol{\beta},\ \sigma_v^2)=f(\boldsymbol{\alpha} \mid \boldsymbol{H},\ \sigma_v^2)$：给定 $\boldsymbol{H}$，$\ln h_t$ 服从 AR(1)模型. 因此，在前

面两部分讨论的 AR 模型的结果可以在这里应用. 特别地，如果$\boldsymbol{\alpha}$的先验分布是均值为$\boldsymbol{\alpha}_o$，协方差阵为C_o的多元正态，则$f(\boldsymbol{\alpha} \mid \boldsymbol{H}, \sigma_v^2)$是均值为$\boldsymbol{\alpha}_*$，协方差阵为$\boldsymbol{C}_*$的多元正态，这里

$$\boldsymbol{C}_*^{-1} = \frac{\sum_{t=2}^{n} \boldsymbol{z}_t \boldsymbol{z}_t'}{\sigma_v^2} + \boldsymbol{C}_o^{-1}, \quad \boldsymbol{\alpha}_* = \boldsymbol{C}_* \left(\frac{\sum_{t=2}^{n} \boldsymbol{z}_t \ln h_t}{\sigma_v^2} + \boldsymbol{C}_o^{-1} \boldsymbol{\alpha}_o \right),$$

其中$\boldsymbol{z}_t = (1, \ln h_{t-1})'$.

- $f(\sigma_v^2 \mid \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{H}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = f(\sigma_v^2 \mid \boldsymbol{H}, \boldsymbol{\alpha})$：给定$\boldsymbol{H}$和$\boldsymbol{\alpha}$，我们可以计算$v_t = \ln h_t - \alpha_0 - \alpha_1 \ln h_{t-1}$ $(t=2, \cdots, n)$. 因此，如果σ_v^2的先验分布是$(m\lambda)/\sigma_v^2 \sim \chi_m^2$，则$\sigma_v^2$的条件后验分布就是一个自由度为$m+n-1$的逆$\chi^2$分布，即，

$$\frac{m\lambda + \sum_{t=2}^{n} v_t^2}{\sigma_v^2} \sim \chi_{m+n-1}^2.$$

注释： 公式(10.23)对$1<t<n$，其中n是样本量. 对于两端的数据点h_1和h_n，需要一些修正. 一个简单的方法是假设h_1是固定的，这样对h_t的抽样就从$t=2$开始. 对于$t=n$，我们用结果$\ln h_n \sim (\alpha_0 + \alpha_1 \ln h_{n-1}, \sigma_v^2)$. 或者，我们可以用$h_{n+1}$的预测和对$h_0$的反向预测，并且继续应用公式. 因为$h_n$是所关心的变量，我们运用在预测原点$n-1$的 2 步向前预测来预测$h_{n+1}$. 对于式(10.21)中的模型，$h_{n+1}$的预测就是

$$\hat{h}_{n-1}(2) = \alpha_0 + \alpha_1(\alpha_0 + \alpha_1 \ln h_{n-1})$$

对h_0的反向预测是基于模型的时间可逆性

$$(\ln h_t - \eta) = \alpha_1(\ln h_{t-1} - \eta) + v_t$$

其中$\eta = \alpha_0/(1-\alpha_1)$和$|\alpha_1|<1$. 反向序列的模型是

$$(\ln h_t - \eta) = \alpha_1(\ln h_{t+1} - \eta) + v_t^*$$

其中$\{v_t^*\}$也是一个零均值，方差为σ_v^2的高斯白噪声序列. 于是，在$t=2$点对h_0的 2 步反向预测就是

$$\hat{h}_2(-2) = \alpha_1^2(\ln h_2 - \eta)$$

420

注释： 公式(10.23)也可以通过使用在 AR(1)模型的一个缺失值的结果来得到；参考 10.6.1 节. 特别地，假设$\ln h_t$是缺失的. 对于式(10.21)中的 AR(1)模型，这个缺失值是与$\ln h_{t-1}$和$\ln h_{t+1}$相关的$(1<t<n)$. 由模型，我们有

$$\ln h_t = \alpha_0 + \alpha_1 \ln h_{t-1} + a_t.$$

定义$y_t = \alpha_0 + \alpha_1 \ln h_{t-1}$，$x_t = 1$和$b_t = -a_t$，我们得到

$$y_t = x_t \ln h_t + b_t \tag{10.24}$$

下面，从

$$\ln h_{t+1} = \alpha_0 + \alpha_1 \ln h_t + a_{t+1}$$

我们定义$y_{t+1} = \ln h_{t+1} - \alpha_0$，$x_{t+1} = \alpha_1$和$b_{t+1} = a_{t+1}$，得到

$$y_{t+1} = x_{t+1}\ln h_t + b_{t+1} \tag{10.25}$$

式(10.24)和式(10.25)构成一个特别简单的含两个观测值和一个未知参数 $\ln h_t$ 的线性回归. 注意到 b_t 和 b_{t+1} 有相同的分布，因为 $-a_t$ 也服从 $N(0, \sigma_v^2)$. $\ln h_t$ 的最小二乘估计就是

$$\widehat{\ln h_t} = \frac{x_t y_t + x_{t+1} y_{t+1}}{x_t^2 + x_{t+1}^2} = \frac{\alpha_0(1-\alpha_1) + \alpha_1(\ln h_{t+1} + \ln h_{t-1})}{1+\alpha_1^2}$$

它正好是式(10.23)中 $\ln h_t$ 的条件均值. 而且，这个估计服从正态分布，均值是 $\ln h_t$，方差是 $\sigma_v^2/(1+\alpha_1^2)$. 公式(10.23)就是 $a_t \sim N(0, h_t)$；和 $\widehat{\ln h_t} \sim N[\ln h_t, \sigma_v^2/(1+\alpha_1^2)]$ 的简单相乘，再加上一个变换 $d\ln h_t = h_t^{-1} dh_t$. 这个回归的方法可以将 $\ln h_t$ 容易地推广到其他 AR(p)模型. 我们使用这个方法，并且假设 $\{h_t\}_{t=1}^{p}$ 对一个随机波动率 AR(p)模型是固定的.

421

注释：h_t 的初始值可以通过对收益率序列拟合第 3 章的波动率模型来得到.

例 10.3 考虑从 1962 年 1 月到 1999 年 12 月的 456 个标准普尔 500 指数的每月对数收益率数据. 图 10-3 描出了根据百分比度量的收益率依时间变化的图形. 如果 GARCH 模型适用于这些数据序列，我们得到一个高斯 GARCH(1，1)模型

$$r_t = 0.658 + a_t, \quad a_t = \sqrt{h_t}\epsilon_t$$

$$h_t = 3.349 + 0.086 a_{t-1}^2 + 0.735 h_{t-1} \tag{10.26}$$

其中系数的 t 比全部都大于 2.52. 标准化残差和他们的平方序列的 Ljung-Box 统计量不能表明模型有不合适之处.

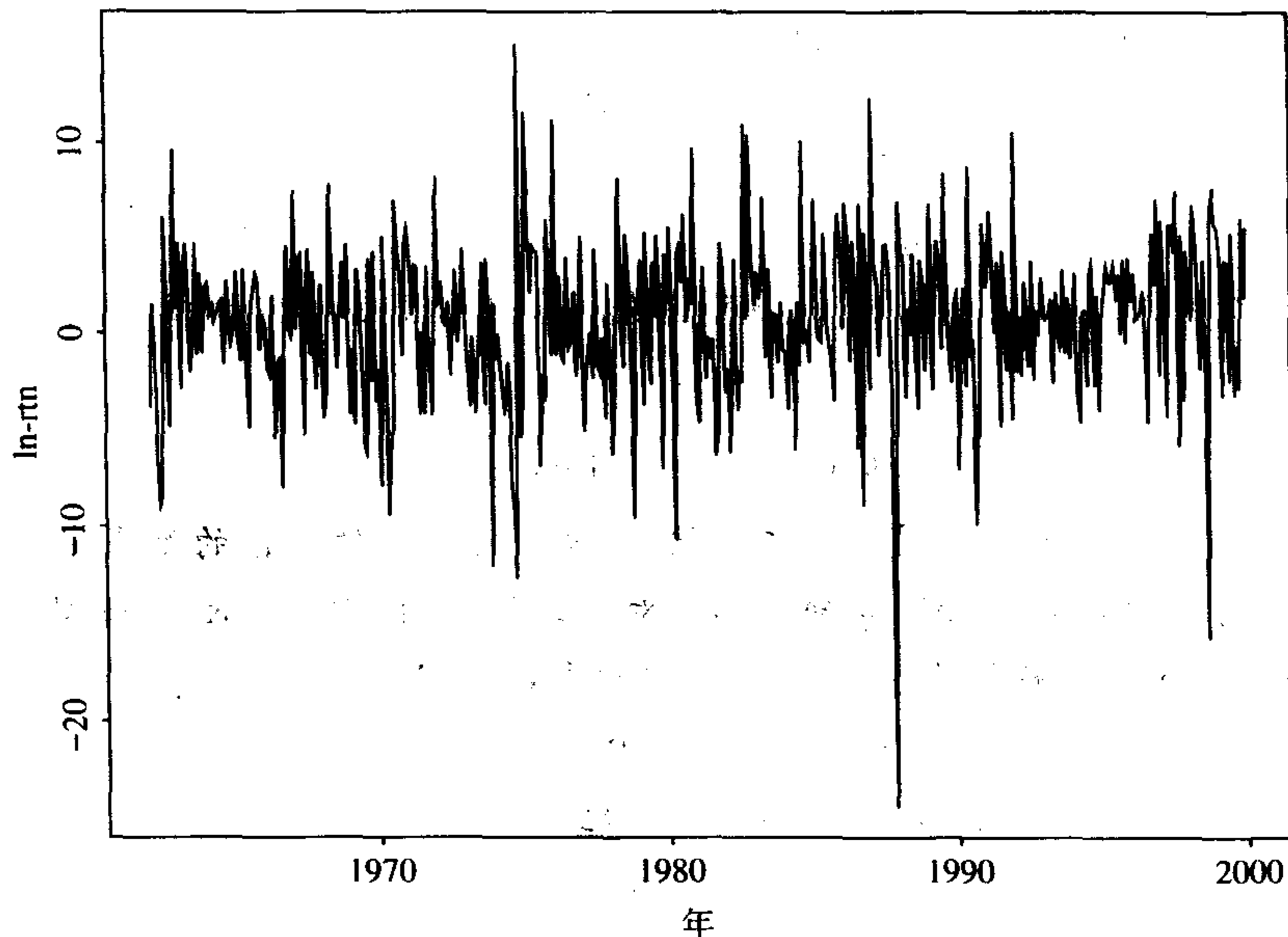


图 10-3 1962 年到 1999 年的标准普尔 500 指数的每月收益率

下面，考虑随机波动率模型

$$r_t = \mu + a_t, a_t = \sqrt{h_t}\varepsilon_t$$
$$\ln h_t = \alpha_0 + \alpha_1 \ln h_{t-1} + v_t \tag{10.27}$$

其中 v_t 是独立同分布的 $N(0, \sigma_v^2)$. 为了应用吉布斯抽样，我们使用先验分布

$$\mu \sim N(0,9), \boldsymbol{\alpha} \sim N[\boldsymbol{\alpha}_o, \text{diag}(0.09, 0.04)], \quad \frac{5 \times 0.2}{\sigma_v^2} \sim \chi_5^2$$

422

其中$\boldsymbol{\alpha}_o=(0.4, 0.8)'$. 对于初始参数值，我们使用对$\{h_t\}$序列拟合的式(10.26)中的GRACH(1，1)模型中的拟合值，并且令 $\sigma_v^2=0.5$ 和 $\mu=0.06$，后者来自样本均值. 并且，h_t 是用格子吉布斯抽样方法使用500个格点抽样得到的，h_t 的范围是$(0, 1.5s^2)$，其中 s^2 是对数收益率 r_t 的样本方差.

我们进行5 100次吉布斯抽样迭代，并且丢掉最初的100次抽样结果. 图10-4所示的是四个系数参数的先验和后验密度函数分布. 先验分布相对无信息先验. 后验分布特别集中在 μ 和 σ_v^2. 图10-5所示的是拟合波动率的依时间变化图形. 上面的一张描出了对每个时间点的5 000次迭代中 h_t 的后验均值，而下面一张描出了式(10.26)中GARCH(1，1)模型的拟合值. 两张图表现出一个相似的图案.

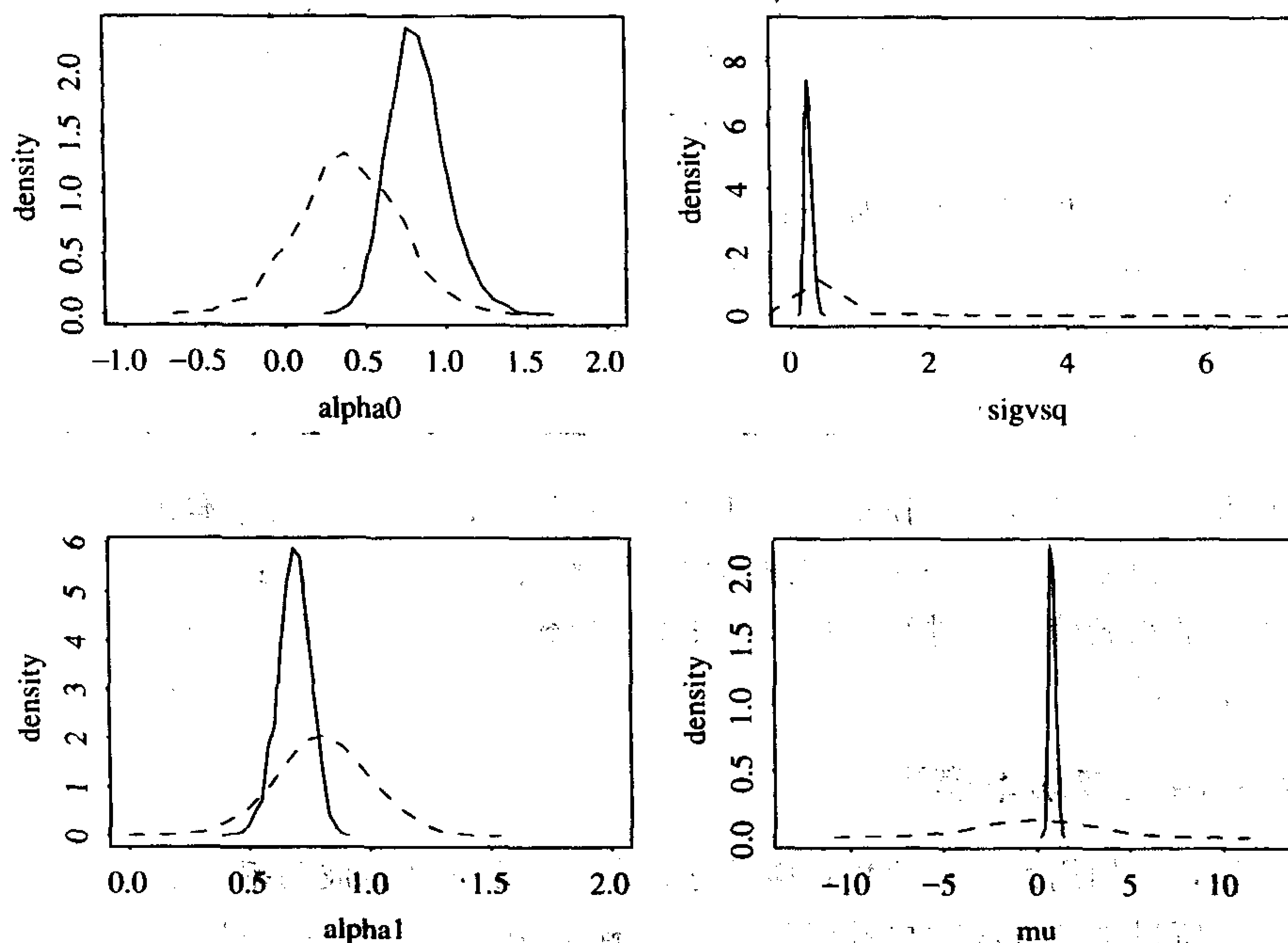


图10-4 对标准普尔500指数的每月对数收益率数据使用的随机波动率模型参数的先验和后验分布密度函数；虚线表示先验密度，实线表示后验密度，它们由5 000次吉布斯迭代抽样得到. 详细参考文字部分

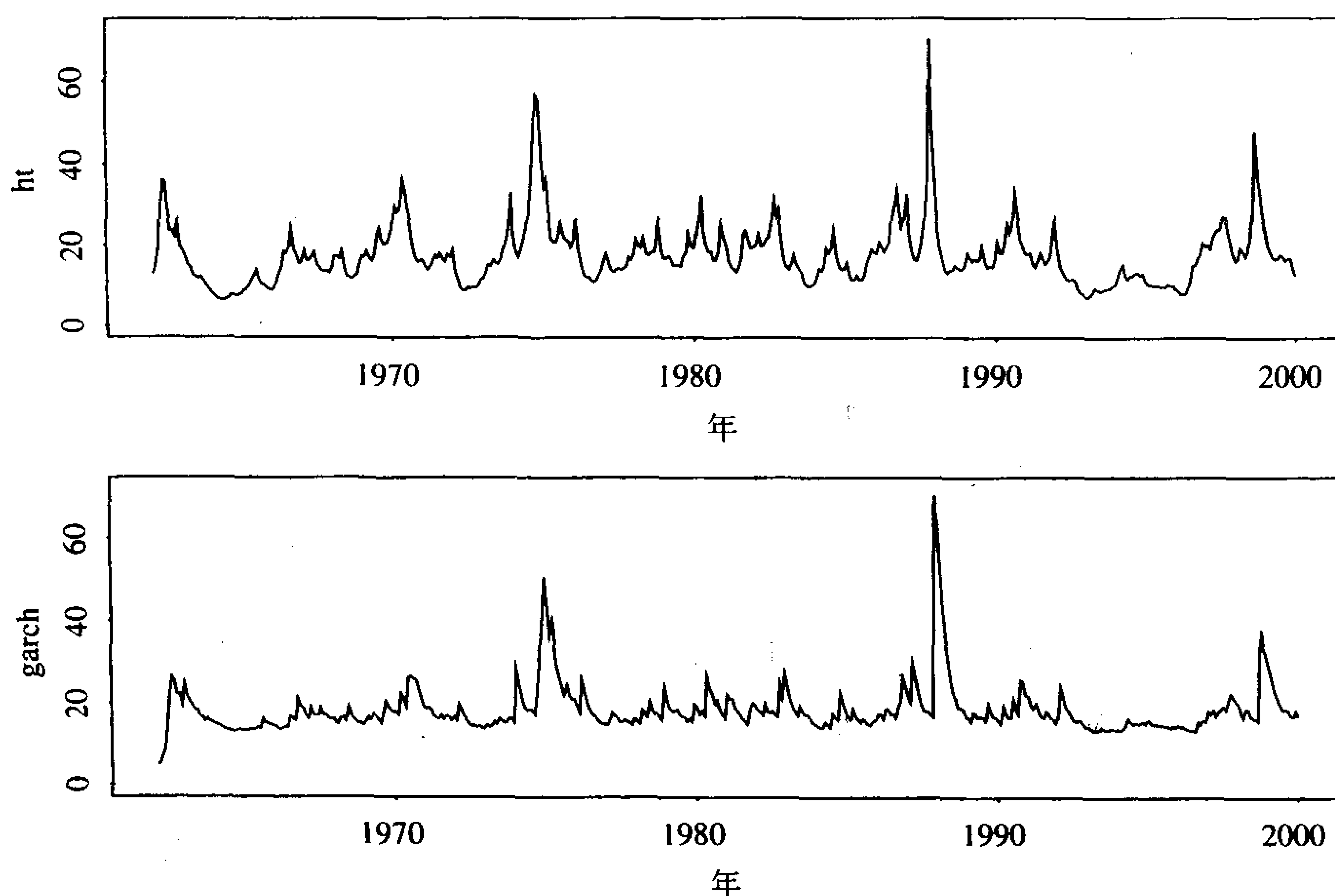


图 10-5 从 1962 年到 1999 年对标准普尔 500 指数的每月对数收益率拟合的波动率依时间变化图．上面的一张描出了 5 000 次迭代的后验均值，下面一张描出了 GARCH(1，1)模型的结果

四个系数的后验均值和标准差如下：

参数	μ	α_0	α_1	σ_v^2
均值	0.836	0.831	0.685	0.265
标准差	0.177	0.183	0.069	0.056

α_1 的后验均值是 0.685，比 Jacquier，Polson 和 Rossi(1994)使用标准普尔 500 指数的日收益率得到的后验均值要小．但是它证实波动率序列的强序列依赖性．最后，我们对其他吉布斯抽样器使用不同的初始值，并进行 3 100 次迭代．参数的后验均值有一点点变化，但是 h_t 的后验均值序列是平稳的．

10.7.2 多元随机波动率模型

在这一节里，我们用第 9 章的楚列斯基分解来研究多元随机波动率模型．我们集中研究二元的情形，但是所讨论的方法对更高维数情形同样适用．基于楚列斯基分解，收益率 $\boldsymbol{r}_t$ 的新息 $\boldsymbol{a}_t$ 变换为 $\boldsymbol{b}_t$，使得

$$b_{1t}=a_{1t},\quad b_{2t}=a_{2t}-q_{21,t}b_{1t},$$

其中 b_{2t} 和 $q_{21,t}$ 可以解释为下面线性回归模型的残差和最小二乘估计

$$a_{2t}=q_{21,t}a_{1t}+b_{2t}.$$

423～424

$\boldsymbol{a}_t$ 的条件协方差矩阵被$\{g_{11,t}, g_{22,t}\}$和$\{q_{21,t}\}$参数化为

$$\begin{bmatrix}\sigma_{11,t} & \sigma_{12,t}\\ \sigma_{21,t} & \sigma_{22,t}\end{bmatrix}=\begin{bmatrix}1 & 0\\ q_{21,t} & 1\end{bmatrix}\begin{bmatrix}g_{11,t} & 0\\ 0 & g_{22,t}\end{bmatrix}\begin{bmatrix}1 & q_{21,t}\\ 0 & 1\end{bmatrix} \tag{10.28}$$

其中 $g_{ii,t}=\mathrm{Var}(b_{it}\mid F_{t-1})$和 $b_{1t}\perp b_{2t}$. 于是，我们关心的量为 $g_{11,t}$，$g_{22,t}$和 $q_{21,t}$.

一个对于收益率向量 $r_t=(r_{1t}, r_{2t})'$的简单的二元随机波动率模型如下：

$$\boldsymbol{r}_t=\boldsymbol{\beta}_0+\boldsymbol{\beta}_1\boldsymbol{x}_t+\boldsymbol{a}_t \tag{10.29}$$

$$\ln g_{ii,t}=\alpha_{i0}+\alpha_{i1}\ln g_{ii,t-1}+v_{it}, i=1,2 \tag{10.30}$$

$$q_{21,t}=\gamma_0+\gamma_1 q_{21,t-1}+u_t. \tag{10.31}$$

其中$\{\boldsymbol{a}_t\}$是一列无序列相关性的高斯随机向量，均值为零，条件协方差矩阵Σ_t 由式(10.28)给出，$\boldsymbol{\beta}_0$ 是一个二维的常数向量，$\boldsymbol{x}_t$ 表示解释变量，$\{v_{1t}\}$，$\{v_{2t}\}$和$\{u_t\}$是三个独立的高斯白噪声序列，使得 $\mathrm{Var}(v_{it})=\sigma_w^2$和 $\mathrm{Var}(u_t)=\sigma_u^2$. 再一次，我们在式(10.30)中使用对数变换，使得 $g_{ii,t}$为正.

令 $\boldsymbol{G}_i=(g_{ii,1}, \cdots, g_{ii,n})'$，$\boldsymbol{G}=[\boldsymbol{G}_1, \boldsymbol{G}_2]$ 和 $\boldsymbol{Q}=(q_{21,1}, \cdots, q_{21,n})'$. 式(10.29)～(10.31)中模型的“传统”参数是$\boldsymbol{\beta}=(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)'$，$\boldsymbol{\omega}_i=(\alpha_{i0}, \alpha_{i1}, \sigma_w^2)$，$i=1$，2 和$\boldsymbol{\gamma}=(\gamma_0, \gamma_1, \sigma_u^2)$. 扩充参数为$\boldsymbol{Q}$，$\boldsymbol{G}_1$，$\boldsymbol{G}_2$. 为了用吉布斯抽样法估计这样一个二元随机波动率模型，我们使用前一小节一元模型的结果和另外两个条件后验分布. 特别地，我们可以抽得下面的随机样本

1)$\boldsymbol{\beta}_0$，$\boldsymbol{\beta}_1$ 的逐行样本，使用结果公式(10.22)；

2)$g_{11,t}$使用公式(10.23)，只要将 a_t 换成a_{1t}；

3)$\boldsymbol{\omega}_1$ 使用与一元情形同样的方法，只是将 a_t 换成a_{1t}.

为了抽得$\boldsymbol{\omega}_2$ 和 $g_{22,t}$的随机样本，我们需要计算 b_{2t}. 而这是容易的，因为给定扩充参数向量$\boldsymbol{Q}$后$b_{2t}=a_{2t}-q_{21,t}a_{1t}$. 而且，$b_{2t}$服从正态分布，均值为0，条件方差为 $g_{22,t}$.

剩下只需要考虑条件后验分布

$$f(\bar{\boldsymbol{\omega}}\mid \boldsymbol{Q},\sigma_u^2),\quad f(\sigma_u^2\mid \boldsymbol{Q},\bar{\boldsymbol{\omega}}),\quad f(q_{21,t}\mid \boldsymbol{A},\boldsymbol{G},\boldsymbol{Q}_{-t},\boldsymbol{\gamma}),$$

其中$\bar{\boldsymbol{\omega}}=(\boldsymbol{\gamma}_0, \gamma_1)'$是式(10.31)中的系数向量，并且 A 表示a_t 的集合，当$\boldsymbol{R}$，$\boldsymbol{X}$，$\boldsymbol{\beta}_0$ 和β_1 给定的时候，它是已知的. 给定$\boldsymbol{Q}$ 和σ_u^2，模型(10.31)是一个简单的高斯 AR(1)模型. 于是，如果$\bar{\boldsymbol{\omega}}$ 的先验分布是均值为$\bar{\boldsymbol{\omega}}_o$、协方差阵为$\boldsymbol{D}_o$的二元正态分布的话，那么$\bar{\boldsymbol{\omega}}$ 的条件后验分布也是二元正态的，均值为$\bar{\boldsymbol{\omega}}_*$，协方差阵为$\boldsymbol{D}_*$. 这里 [425]

$$\boldsymbol{D}_*^{-1}=\frac{\sum_{t=2}^{n} z_t z_t'}{\sigma_u^2}+\boldsymbol{D}_o^{-1},\quad \bar{\boldsymbol{\omega}}_*=\boldsymbol{D}_*\left(\frac{\sum_{t=2}^{n} z_t q_{21,t}}{\sigma_u^2}+\boldsymbol{D}_o^{-1}\bar{\boldsymbol{\omega}}_o\right)$$

其中 $z_t=(1, q_{21,t-1})'$. 同样地，如果 σ_u^2 的先验分布为$(m\lambda)/\sigma_u^2\sim\chi_m^2$，那么 σ_u^2 的条件后验分布为

$$\frac{m\lambda+\sum_{t=2}^{n}u_t^2}{\sigma_u^2}\sim\chi_{m+n-1}^2$$

其中 $u_t=q_{21,t}-\gamma_0-\gamma_1 q_{21,t-1}$. 最后,

$$
\begin{aligned}
& f(q_{21,t} \mid \boldsymbol{A},\boldsymbol{G},\boldsymbol{Q}_{-t},\sigma_u^2,\bar{\boldsymbol{\omega}}) \qquad (10.32)\\
\propto & f(b_{2t} \mid g_{22,t}) f(q_{21,t} \mid q_{21,t-1},\bar{\boldsymbol{\omega}},\sigma_u^2) f(q_{21,t+1} \mid q_{21,t},\bar{\boldsymbol{\omega}},\sigma_u^2)\\
\propto & g_{22,t}^{-0.5}\exp[-(a_{2t}-q_{21,t}a_{1t})^2/(2g_{22,t})]\exp[-(q_{21,t}-\mu_t)^2/(2\sigma^2)]
\end{aligned}
$$

其中 $\mu_t=[\gamma_0(1-\gamma_1)+\gamma_1(q_{21,t-1}+q_{21,t+1})]/(1+\gamma_1^2)$ 和 $\sigma^2=\sigma_u^2/(1+\gamma_1^2)$. 一般地, μ_t 和 σ^2 可以通过使用在 AR(p)过程中的缺失值的结果来得到. 结果是方程(10.32)有一个关于 $q_{21,t}$ 分布的显式表达式. 特别地, 式(10.32)的第一项, 即给定 $g_{22,t}$ 和 a_t 条件下, $q_{21,t}$ 的条件分布, 是正态的, 均值为 a_{2t}/a_{1t}, 方差为 $g_{22,t}/a_{1t}^2$. 该式的第二项也是正态的, 均值为 μ_t, 方差为 σ^2. 于是, 由 10.3 节的结果 1, $q_{21,t}$ 的条件后验分布是正态的, 均值为 μ_*, 方差为 σ_*^2, 其中

$$
\frac{1}{\sigma_*^2}=\frac{a_{1t}^2}{g_{22,t}}+\frac{1+\gamma_1^2}{\sigma_u^2}, \quad \mu_*=\sigma_*^2\left(\frac{1+\gamma_1^2}{\sigma_u^2}\times\mu_t+\frac{a_{1t}^2}{g_{22,t}}\times\frac{a_{2t}}{a_{1t}}\right)
$$

而 μ_t 由式(10.32)定义.

例 10.4 在这个例子里, 我们研究从 1962 年 1 月到 1999 年 12 月的 IBM 股票和标准普尔 500 指数的每月对数收益率数据的二元波动率模型. 这是例 10.3 的扩充, 加入了 IBM 股票收益率. 图 10-6 所示这两个收益率序列的时间图. 令 $\boldsymbol{r}_t=(\text{IBM}_t, \text{SP}_t)'$. 如果第 9 章中的经过楚列斯基分解的时变相关 GARCH 模型适用的话, 我们得到模型

426

$$\boldsymbol{r}_t=\boldsymbol{\beta}_0+\boldsymbol{a}_t \qquad (10.33)$$

$$g_{11,t}=\alpha_{10}+\alpha_{11}g_{11,t-1}+\alpha_{12}a_{1,t-1}^2 \qquad (10.34)$$

$$g_{22,t}=\alpha_{20}+\alpha_{21}a_{1,t-1}^2 \qquad (10.35)$$

$$q_{21,t}=\gamma_0 \qquad (10.36)$$

其中估计及它们的标准差由表 10-2(a)给出. 为了比较, 我们使用与方程(10.33)同样的均值方程和一个与方程(10.34)~(10.36)相似的随机波动率模型. 波动率方程是

$$\ln g_{11,t}=\alpha_{10}+\alpha_{11}\ln g_{11,t-1}+v_{1t}, \quad \text{Var}(v_{1t})=\sigma_{1v}^2 \qquad (10.37)$$

$$\ln g_{22,t}=\alpha_{20}+v_{2t}, \quad \text{Var}(v_{2t})=\sigma_{2v}^2 \qquad (10.38)$$

$$q_{21,t}=\gamma_0+u_t, \quad \text{Var}(u_t)=\sigma_u^2 \qquad (10.39)$$

使用的先验分布是

$$
\beta_{i0}\sim N(0.8,4), \quad \alpha_1\sim N[(0.4,0.8)',\text{diag}(0.16,0.04)], \quad \alpha_{20}\sim N(5,25)
$$

$$
\gamma_0\sim N(0.4,0.04), \quad \frac{10\times 0.1}{\sigma_{1v}^2}\sim\chi_{10}^2, \quad \frac{5\times 0.2}{\sigma_{2v}^2}\sim\chi_5^2, \quad \frac{5\times 0.2}{\sigma_u^2}\sim\chi_5^2
$$

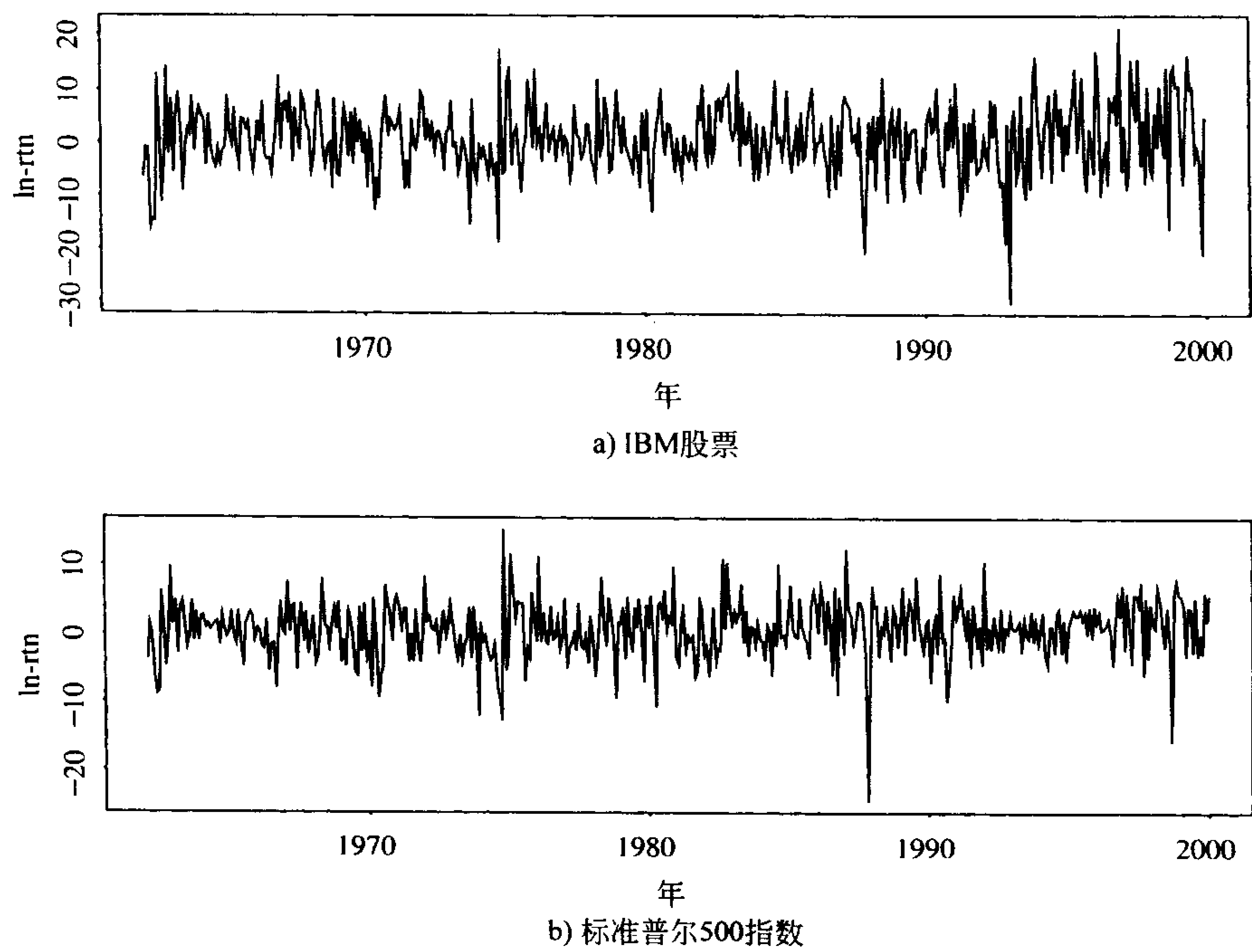


图 10-6　从 1962 年到 1999 年的 IBM 股票和标准普尔 500 指数的每月对数收益率时间图

427

表　10-2

(a)带时变相关系数的二元 GARCH(1，1)模型

参数	β_{01}	β_{02}	α_{10}	α_{11}	α_{12}	α_{20}	α_{21}	γ_0
估计	1.04	0.79	3.16	0.83	0.10	10.59	0.04	0.35
标准差	0.31	0.20	1.67	0.08	0.03	0.93	0.02	0.02

(b)随机波动率模型

参数	β_{01}	β_{02}	α_{10}	α_{11}	σ^2_{1v}	α_{20}	σ^2_{2v}	γ_0	σ^2_u
后验均值	0.86	0.84	0.52	0.86	0.08	1.81	0.39	0.39	0.08
标准差	0.30	0.18	0.18	0.05	0.03	0.11	0.06	0.03	0.02

注：从 1962 年 1 月到 1999 年 12 月 IBM 股票和标准普尔 500 指数的每月对数收益率数据的二元波动率模型的估计．随机波动率模型基于总数为 1 300 次的最后 1 000 次吉布斯抽样迭代．

这些先验分布是相对无信息的．我们进行吉布斯抽样 1 300 次迭代，但是丢掉前 300 次的迭代结果．$g_{ii,t}$ 的随机样本是用 400 个格点的格子吉布斯抽样抽得的，间隔为$[0,\ 1.5s_i^2]$，其中 s_i^2 是对数收益率 r_{it} 的样本方差．表 10-2(b)给出了二元随机波动率模型的“传统”参数的后验均值和标准差．

为了检测吉布斯抽样的收敛性，我们运行此过程若干次，每次用不同的初始值和不同的迭代次数，结果是稳定的．图 10-7 描出了两个不同吉布斯抽样过程的各种量的散点图．第一个吉布斯抽样是基于 300＋1 000 次迭代的，第二个吉布斯抽样是基于 500＋3 000 次迭代的，其中 $M+N$ 表示一共进行 $M+N$ 次迭代，但是丢掉前 M 次迭代的结果．散点图分别对应于

$g_{11,t}$，$g_{22,t}$，$q_{21,t}$，$\sigma_{22,t}$，$\sigma_{21,t}$的后验均值和相关系数$\rho_{21,t}$. 在每张图上添加的直线 $y=x$ 是为了表现后验均值的相近程度. 吉布斯抽样结果的稳定性是显然的.

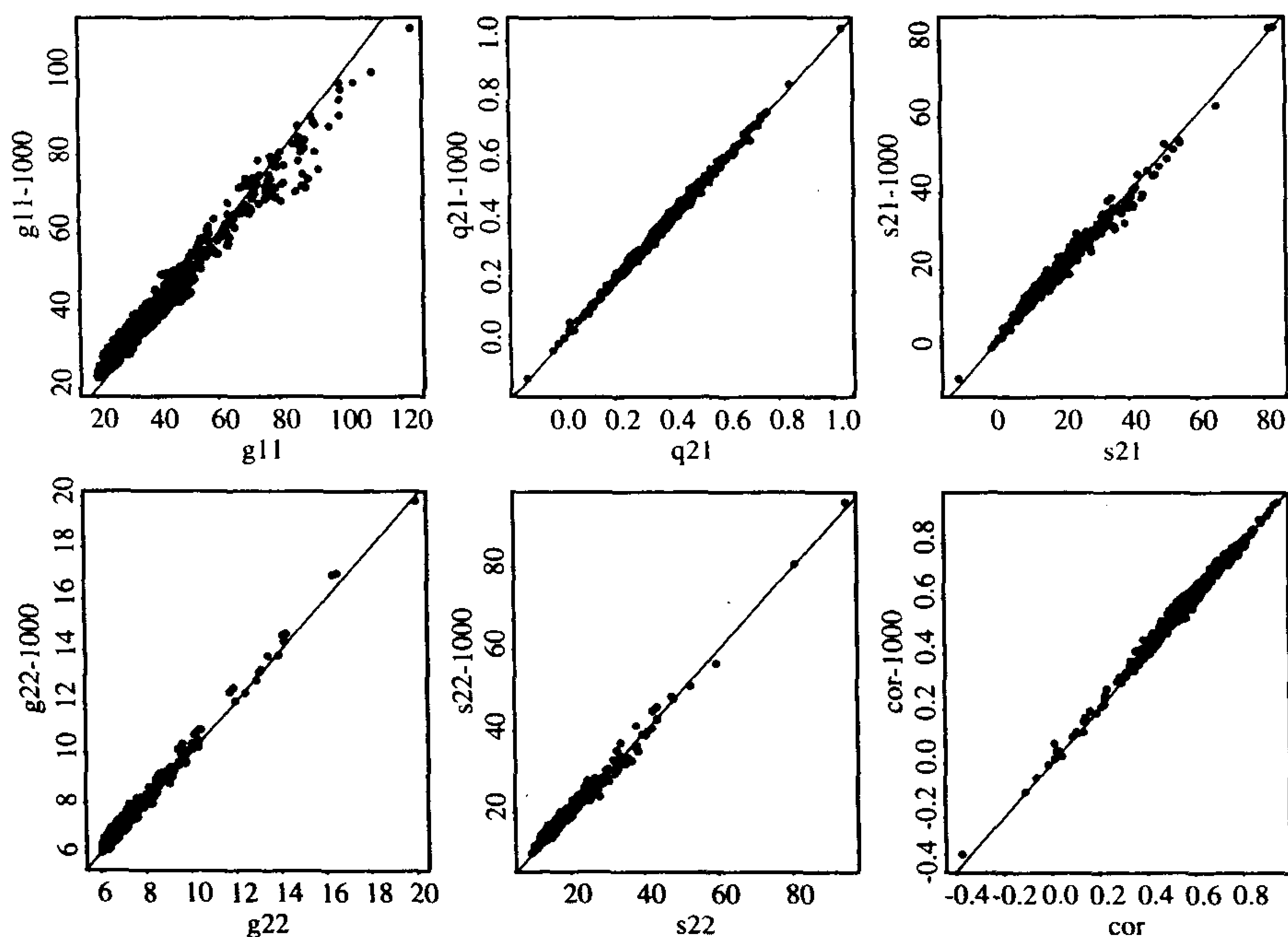
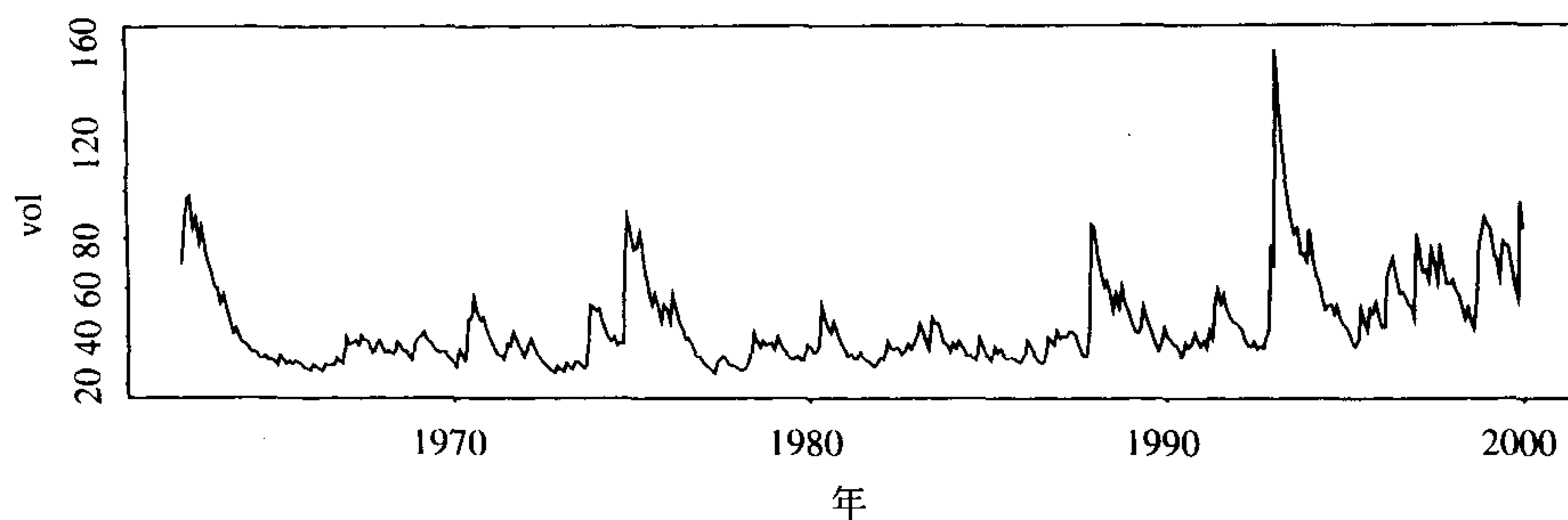
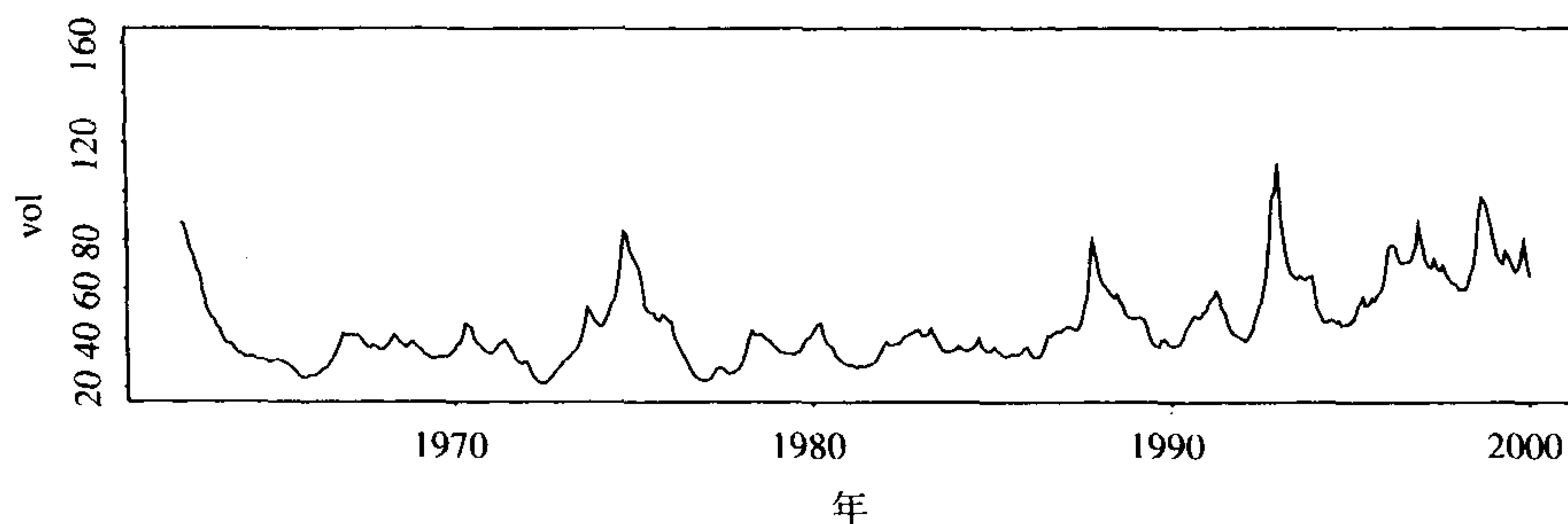


图 10-7 IBM 股票和标准普尔 500 指数的每月对数收益率数据的二元随机波动率模型的两种吉布斯抽样的各种统计量的后验均值的散点图. x 轴表示基于 500+3 000 次迭代的结果，y 轴表示基于 300+1 000 次迭代的结果. 记号如文中所述

比较一下带有方程组(10.33)～(10.36)中时变相关系数的 GARCH 模型和随机波动率模型是有意思的. 首先，正如预期的，此两个模型的均值方程在本质上是一样的. 第二，图 10-8描出了 IBM 股票收益率的拟合波动率的时间图. 上面一张是对应于 GARCH 模型的，下面一张描出了随机波动率模型的后验均值. 这两个模型呈现相似的波动率特征；他们展示了波动率聚类，还表明在波动率中有一个上升的趋势. 然而，GARCH 模型生成更高的波动率峰值. 第三，图 10-9 描出了标准普尔 500 指数收益率的拟合波动率的时间图. GARCH 模型在 1993 年附近产生了一个额外的波动率峰值. 这个额外的峰值没有在图 10-5 展示的一元分析中出现. 看上去对于这个特殊的例子，二元 GARCH 模型生成的是一个假的波动率峰值. 这个假的峰值是由对 IBM 收益率的依赖性引起的，而且并没有在随机波动率模型中出现. 实际上，标准普尔 500 指数收益率用二元随机波动率模型得出的拟合波动率与一元分析得出的相似. 第四，图 10-10 描出了拟合条件相关系数的时间图. 由此表明两个模型中有本质区别. GARCH 模型的相关系数相对平滑而且是正的，有均值 0.55 和标准差 0.11. 然而，由随机波动率模型产生的相关系数在不同的月份有显著的不同，均值为 0.57、标准差为 0.17. 而且，在一些特殊情况下相关系数还是负数. 这个差异是可以理解的，因为在随机波动率模型中，$q_{21,t}$包含了随机抖动 u_t.

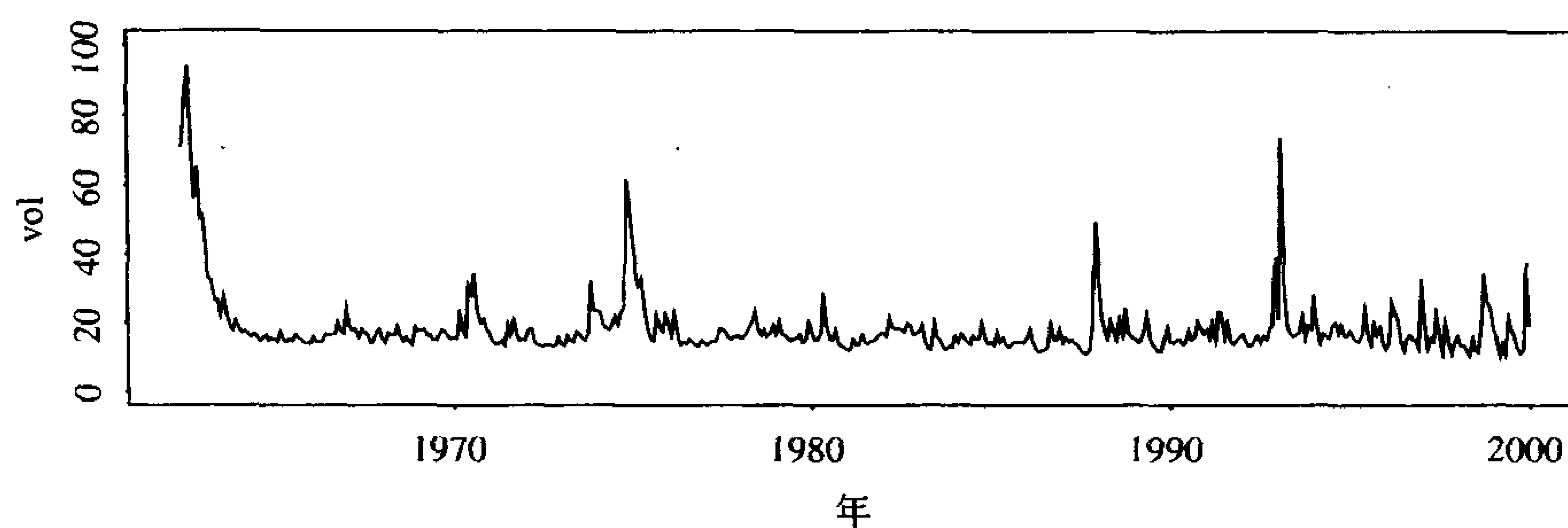


a) 带时变自相关系数的GARCH模型

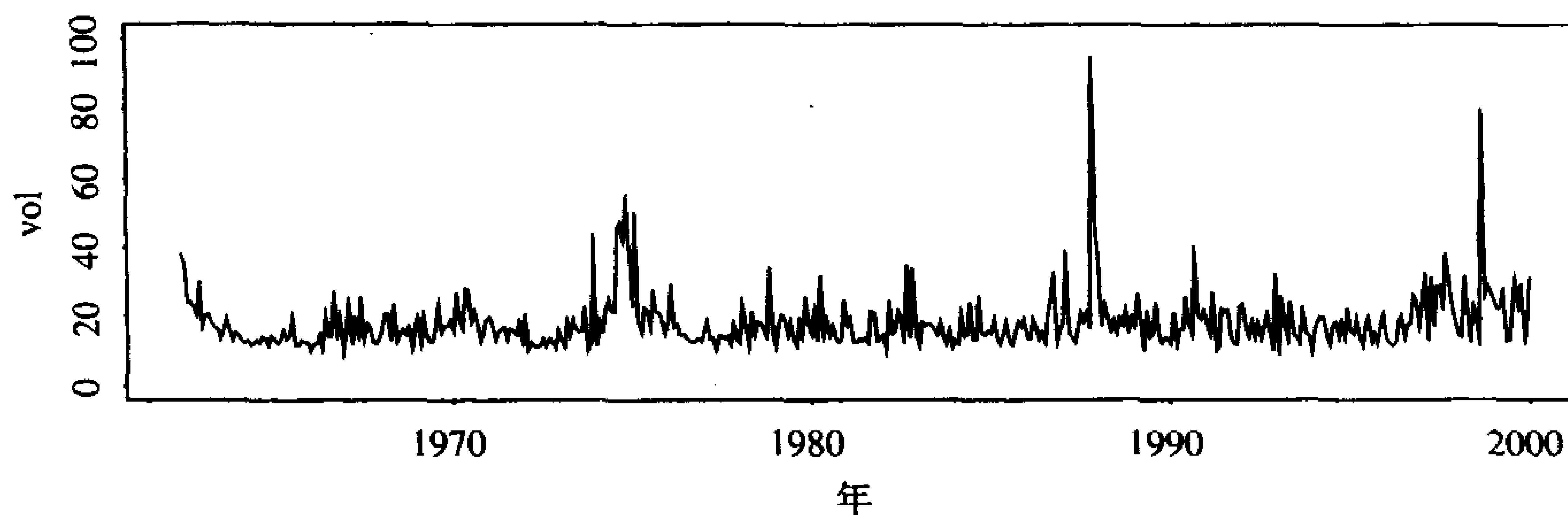


b) 运用吉布斯抽样器重复迭代300+1 000次估计出来的二元随机波动率模型

图 10-8 从 1962 年到 1999 年 IBM 股票每月对数收益率数据的拟合波动率时间图



a) 带时变自相关系数的GARCH模型



b) 运用吉布斯抽样器重复迭代300+1000次估计出来的随机波动率模型

图 10-9 从 1962 年到 1999 年的标准普尔 500 指数每月对数收益率数据的拟合波动率时间图

注释： 吉布斯抽样估计法对于其他二元随机波动率模型同样适用，但条件后验分布是这一节讨论中的条件后验分布的推广，不过它们基于相同的想法.

10.8 马尔可夫转换模型

马尔可夫转换模型是又一个用 MCMC 方法比用其他传统似然方法有更多优势的计量经济学模型. McCulloch 和 Tsay (1994)讨论了一个吉布斯抽样过程，去估计在每一个状态波动率都不随时间变化的模型，并把他们的方法用于估计美国的季度国民生产总值增长率(已做季节性调整)，对不同的状态有不同的动态机制和均值水平的马尔可夫转换模型，得到一些有趣的结果. 例如，增长率的动态变化在不同的经济“扩张”和“收缩”期之间有明显的不同. 由于这一章关心的是资产收益率，我们集中关注模型的波动率转换.

假定资产收益率 r_t 服从一个简单的两个状态的转换模型，该模型带有不同风险溢价和不同的 GARCH 动态机制：

$$r_t = \begin{cases} \beta_1\sqrt{h_t} + \sqrt{h_t}\epsilon_t, & h_t = \alpha_{10} + \alpha_{11}h_{t-1} + \alpha_{12}a_{t-1}^2 \quad 若\ s_t = 1 \\ \beta_2\sqrt{h_t} + \sqrt{h_t}\epsilon_t, & h_t = \alpha_{20} + \alpha_{21}h_{t-1} + \alpha_{22}a_{t-1}^2 \quad 若\ s_t = 2 \end{cases} \tag{10.40}$$

其中 $a_t=\sqrt{h_t}\epsilon_t$，$\{\epsilon_t\}$是均值为 0、方差为 1 的高斯白噪声序列，参数 α_{ij} 满足一些正规性条件，使得 a_t 的无条件方差存在. 从一个状态到另一个状态的概率转移由下式控制

$$P(s_t = 2 \mid s_{t-1} = 1) = e_1, \quad P(s_t = 1 \mid s_{t-1} = 2) = e_2 \tag{10.41}$$

这里 $0<e_i<1$. 一个小 e_i 意味着收益率序列倾向于在同一状态 i 逗留，期望持续期为 $1/e_i$. 对方程(10.40)中的待识别模型，我们假定 $\beta_2>\beta_1$，使得状态 2 与更高的风险溢价相联系. 这并不是一个关键性的限制，因为它用来得到作状态标记的唯一性. 模型的一个特殊情形是对所有的 j，$\alpha_{1j}=\alpha_{2j}$，即对所有的状态都假定一个 GARCH 模型. 然而，如果用 β_i 代替 $\beta_i\sqrt{h_t}$，则模型(10.40)简化为简单的马尔可夫转换 GARCH 模型.

428～431

模型(10.40)是一个马尔可夫转换 GARCH-M 模型. 为了简便，我们假定 h_1 的初始波动率是给定的，其值等于 r_t的样本方差. 一个更加熟练的分析是将 h_1 作为一个参数，并与其他参数联合估计它. 我们期望在大多数应用中，固定 h_1 的效应都可以忽略，尤其当样本量很大时. 马尔可夫转换 GARCH-M 模型的“传统”参数为$\boldsymbol{\beta}=(\beta_1, \beta_2)'$，$\boldsymbol{\alpha}_i=(\alpha_{i0}, \alpha_{i1}, \alpha_{i2})'$，$i=1$ 和 2，转移概率 $\boldsymbol{e}=(e_1, e_2)'$. 状态向量 $\boldsymbol{S}=(s_1, s_2, \cdots, s_n)'$包含扩张参数. 如果给定 h_1，$\boldsymbol{\alpha}_i$ 和状态向量 $\boldsymbol{S}$，则波动率向量 $\boldsymbol{H}=(h_2, \cdots, h_n)'$可以递推地计算.

模型(10.40)中收益率对波动率的依赖蕴含了收益率也是序列相关的. 这样，此模型具有收益率方面的一些可预测性. 然而，将来收益率的状态是未知的，而且由模型产生的预测必须是所有状态配置上那些预测的一个混合. 这通常导致将来收益率的点预测的较高不确定性.

转为考虑估计. 模型(10.40)中的似然函数是复杂的，因为它是所有可能的状态配置的一个混合. 而吉布斯抽样法仅仅要求下面的条件后验分布

$$f(\boldsymbol{\beta} \mid \boldsymbol{R},\boldsymbol{S},\boldsymbol{H},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2), \quad f(\boldsymbol{\alpha}_i \mid \boldsymbol{R},\boldsymbol{S},\boldsymbol{H},\boldsymbol{\alpha}_{j\neq i}),$$

$$P(\boldsymbol{S} \mid \boldsymbol{R}, h_1, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2), \quad f(e_i \mid \boldsymbol{S}), i = 1,2$$

这里 $\boldsymbol{R}$ 是观测收益率的集合. 为了简便，我们利用 10.3 节讨论的共轭先验分布，即

$$\beta_i \sim N(\beta_{io}, \sigma_{io}^2), \quad e_i \sim \text{Beta}(\gamma_{i1}, \gamma_{i2}).$$

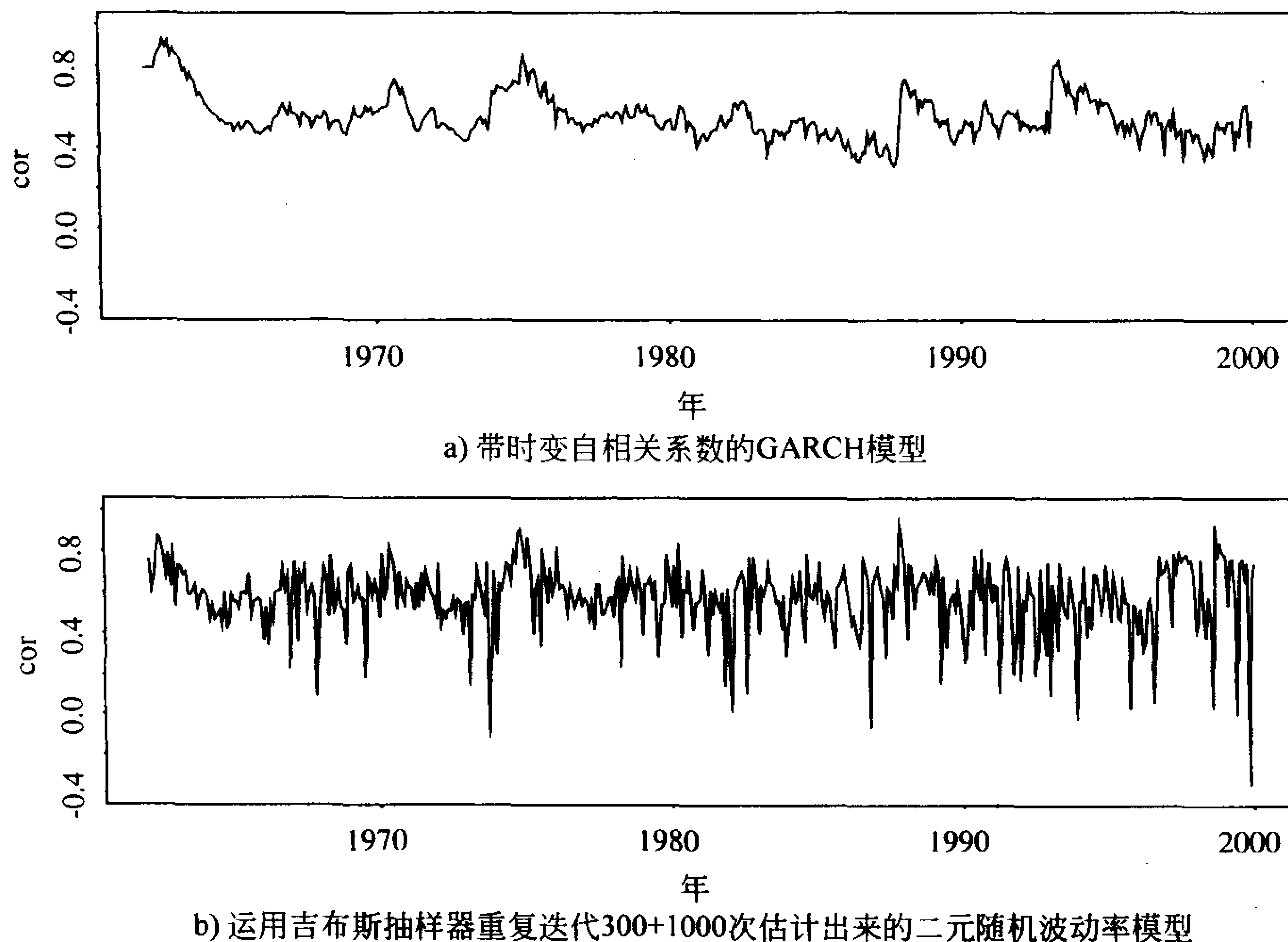


图 10-10 从 1962 年到 1999 年的 IBM 股票和标准普尔 500 指数每月对数收益率数据拟合的相关系数时间图

432

参数 α_{ij} 的先验分布在一个恰当指定的区间上是均匀的. 因为 α_{ij} 是似然函数的非线性参数，我们利用格子吉布斯抽取它的随机实现. 均匀先验分布简化了涉及的计算. 条件后验分布的细节如下给出：

1)β_i 的后验分布仅仅依赖于处于状态 i 中的数据，定义

$$r_{it} = \begin{cases} r_t / \sqrt{h_t} & 若\ s_t = i \\ 0 & 其他 \end{cases},$$

则我们有

$$r_{it} = \beta_i + \varepsilon_t, \quad 对\ s_t = i.$$

因此，数据关于 β_i 的信息包含在 r_{it} 的样本均值之中. 令 $\bar{r}_i = \sum_{s_t = i} r_{it} / n_i$，这里的求和是对处于状态 i 的所有数据点求和，n_i 是处于状态 i 的数据点的数量，则 β_i 的条件后验分布是正态的均值为 β_i^*，方差为 σ_{i*}^2，这里

$$\frac{1}{\sigma_{i*}^2} = n_i + \frac{1}{\sigma_{io}^2}, \quad \beta_i^* = \sigma_{i*}^2 (n_i \bar{r}_i + \beta_{io} / \sigma_{io}^2), \qquad i = 1,2.$$

2)下一步，参数 α_{ij} 可以利用格子吉布斯方法逐个抽取. 给定 h_1，$\boldsymbol{S}$，$\boldsymbol{\alpha}_{v\neq i}$ 和 $\alpha_{iv}(v\neq j)$，则 α_{ij} 的条件后验分布并不对应于一个著名的分布，但是它可以很容易地估计为

$$f(\alpha_{ij} \mid .) \propto -\frac{1}{2}\sum_{s_t=i}\left[\ln h_t + \frac{(r_t-\beta_i\sqrt{h_t})^2}{h_t}\right],$$

其中 h_t 含有 α_{ij}. 我们对一个恰当指定的区间上 α_{ij} 的一列格点估计这个函数. 例如 $0\leqslant\alpha_{11}<1-\alpha_{12}$.

3)e_i 的条件后验分布只涉及 $\boldsymbol{S}$. 令 ℓ_1 是在 $\boldsymbol{S}$ 中从状态 1 到状态 2 转换的数量，ℓ_2 是从状态 2 到状态 1 转换的数量. 再令 n_i 表示处于状态 i 的数据点的数量. 则，由共轭先验分布的结果 3，e_i 的后验分布是贝塔$(\gamma_{i1}+\ell_i, \gamma_{i2}+n_i-\ell_i)$.

4)最后，$\boldsymbol{S}$ 中元素可以逐个抽取. 令 $\boldsymbol{S}_{-j}$ 表示将 $\boldsymbol{S}$ 中去掉 s_j 之后得到的向量. 给定 $\boldsymbol{S}_{-j}$ 与其他信息，可以假定 s_j 有两种可能(即 $s_j=1$ 或 $s_j=2$)，且它的条件后验分布为

433

$$P(s_j \mid .) \propto \prod_{t=j}^{n} f(a_t \mid \boldsymbol{H})P(s_j \mid \boldsymbol{S}_{-j}),$$

概率 $P(s_j=i \mid \boldsymbol{S}_{-j})=P(s_j=i \mid s_{j-1}, s_{j+1})$，$i=1, 2$，可以由方程(10.41)的马尔可夫转移概率计算. 另外，假定 $s_j=i$，则对 $t\geqslant j$，可以递推地计算 h_t. 相关的似然函数用 $L(s_j)$ 表示，它由下式给出，对 $i=1, 2$，

$$L(s_j=i) \equiv \prod_{t=j}^{n} f(a_t \mid \boldsymbol{H}) \propto \exp(f_{ji}), f_{ji} = \sum_{t=j}^{n} -\frac{1}{2}\left[\ln(h_t)+\frac{a_t^2}{h_t}\right].$$

如果 $s_t=1$，则 $a_t=r_t-\beta_1\sqrt{h_t}$；否则，$a_t=r_t-\beta_2\sqrt{h_t}$. 因此，$s_j=1$ 的条件后验概率为

$$P(s_j=1 \mid .) = \frac{P(s_j=1 \mid s_{j-1},s_{j+1})L(s_j=1)}{P(s_j=1 \mid s_{j-1},s_{j+1})L(s_j=1)+P(s_j=2 \mid s_{j-1},s_{j+1})L(s_j=2)}.$$

从而状态 s_j 可以利用单位区间[0, 1]上的均匀分布很容易地抽取.

注释：因为当 e_1 和 e_2 都很小时，s_j 与 s_{j+1} 是高度相关的，所以联合抽取几个 s_j 更加有效. 然而，随着联合抽取状态的数量增加，枚举出的可能状态涉及的计算也快速增加.

例 10.5 本例中，我们考虑 General Electric 公司从 1926 年 1 月至 1999 年 12 月的月对数收益率，共有 888 个观测值. 该收益率是用百分比表示的，并在图 10-11a 中显示. 为了比较，我们对序列以一个 GARCH-M 模型开始，得到

$$r_t = 0.182\sqrt{h_t}+a_t, \quad a_t=\sqrt{h_t}\varepsilon_t,$$
$$h_t = 0.546+1.740h_{t-1}-0.775h_{t-2}+0.025a_{t-1}^2. \tag{10.42}$$

r_t 是月对数收益率，$\{\varepsilon_t\}$是均值为 0、方差为 1 的独立高斯白噪声序列. 所有参数的估计都是高度显著的，其 p 值都小于 0.000 6. 标准化残差及其平方序列的 L-B 统计量没能显示任何的模型不合适性. 这重新保证了风险溢价是正的，而且是显著的. 方程(10.42)中的 GARCH 模型可以写为

434

$$(1-1.765B+0.775B^2)a_t^2 = 0.546+(1-0.025B)\eta_t,$$

其中 $\eta_t=a_t^2-h_t$，B 是向后推移算子，满足 $Ba_t^2=a_{t-1}^2$. 如第 3 章中讨论的，上述方程可以认为

是平方序列 a_t^2 的一个带非齐次新息的 ARMA(2, 1)模型. AR 多项式可以因式分解为$(1-0.945B)(1-0.820B)$，这表明它具有两个小于 1 的实特征根. 因此，r_t 的无条件方差有限，且等于 $0.546/(1-1.765+0.775)\approx49.64$.

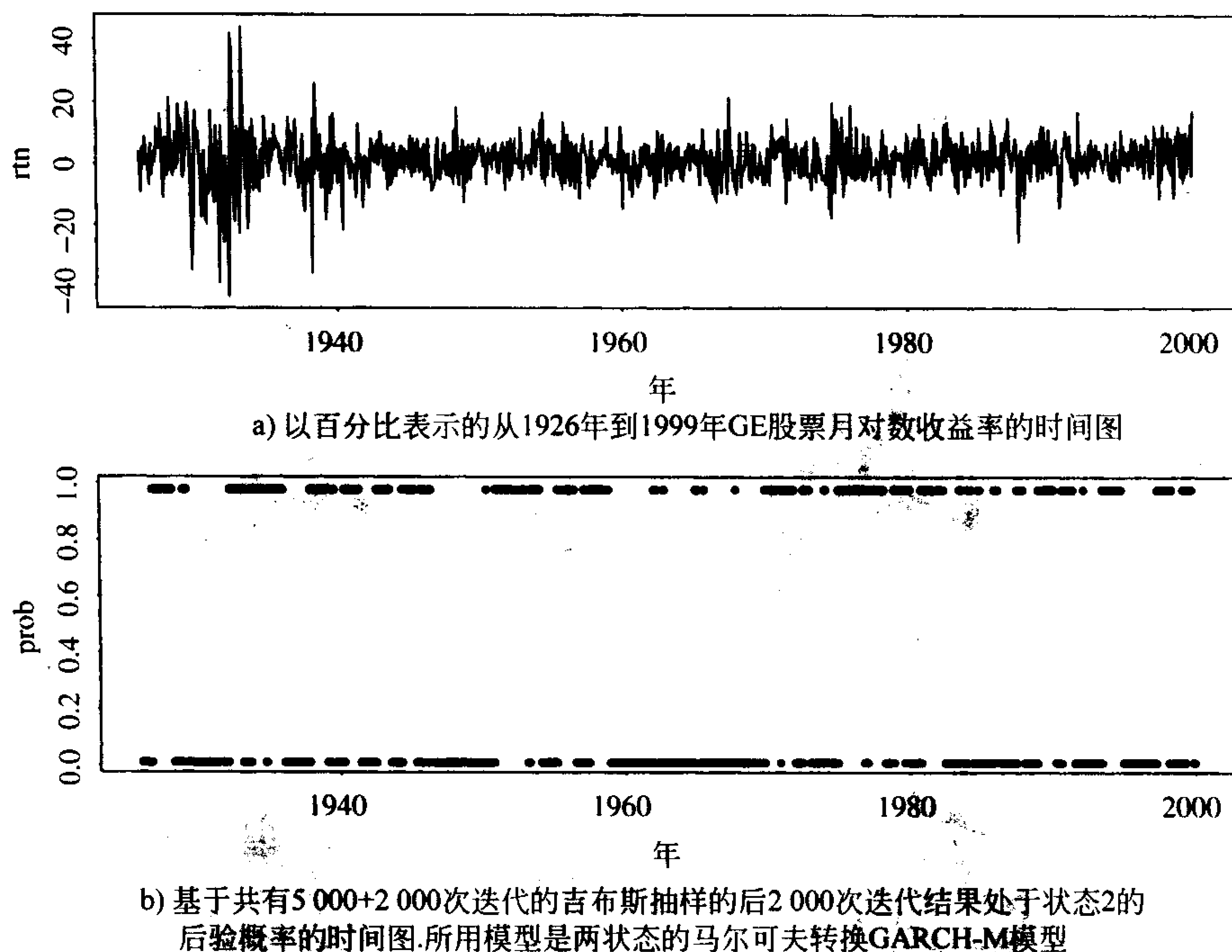
现在来考虑马尔可夫转换模型. 我们利用下面的先验分布：

$$\beta_1 \sim N(0.3,0.09), \quad \beta_2 \sim N(1.3,0.09), \quad \varepsilon_i \sim \text{Beta}(5,95).$$

初始参数值为(a)$e_i=0.1$；(b)s_1 是一个具有等概率的伯努利实验，s_t 是用初始转移概率产生的序列. (c)$\alpha_1=(1.0, 0.6, 0.2)'$，$\alpha_2=(2, 0.7, 0.1)'$. α_{ij} 是用具有 400 个格点的格子吉布斯抽取的，它们在下面范围是等间隔的：

$$\alpha_{i0} \in [0,6.0], \quad \alpha_{i1} \in [0,1], \quad \alpha_{i2} \in [0,0.5].$$

另外，对于 $i=1, 2$，我们实施限制 $\alpha_{i1}+\alpha_{i2}<1$. 吉布斯抽样进行 5 000+2 000 次迭代，但是仅用最后 2 000 次迭代的结果来进行推断.



a) 以百分比表示的从1926年到1999年GE股票月对数收益率的时间图

b) 基于共有5 000+2 000次迭代的吉布斯抽样的后2 000次迭代结果处于状态2的后验概率的时间图.所用模型是两状态的马尔可夫转换GARCH-M模型

图 10-11

435

表 10-3 显示了方程(10.40)中马尔可夫转换 GARCH-M 模型参数的后验均值和标准差. 它也包含了一些统计量. 可以显示两状态之间的差别，如 $\theta=\beta_2-\beta_1$. 风险溢价之差在 5%水平下是统计上显著的. 两状态之间的波动率参数的后验均值之差看上去并不显著. 而波动率参数的后验分布显示了一些不同的特征. 图 10-12 和图 10-13 显示了马尔可夫转换 GARCH-M 模型中所有参数的直方图，显示了两状态之间的某些差别，图 10-14 显示了两状态的持续参数

$\alpha_{i1}+\alpha_{i2}$的时间图，它说明了状态 1 的持续参数频繁地到达边界 1.0，但是状态 2 不是这样．两状态的期望持续期分别是 11 和 9 个月．图 10-11b 显示了每个观测处在状态 2 的后验概率．

表 10-3

(a)状态 1					
参数	β_1	e_1	α_{10}	α_{11}	α_{12}
后验均值	0.111	0.089	2.070	0.844	0.033
后验标准差	0.043	0.012	1.001	0.038	0.033
(b)状态 2					
参数	β_2	e_2	α_{20}	α_{21}	α_{22}
后验均值	0.247	0.112	2.740	0.869	0.068
后验标准差	0.050	0.014	1.073	0.031	0.024
状态之间的差					
参数	$\beta_2-\beta_1$	e_2-e_1	$\alpha_{20}-\alpha_{10}$	$\alpha_{21}-\alpha_{11}$	$\alpha_{22}-\alpha_{12}$
后验均值	0.135	0.023	0.670	0.026	−0.064
后验标准差	0.063	0.019	1.608	0.050	0.043

注：对 1926 年 1 月到 1999 年 12 月 GE 股票的月对数收益率拟合的马尔可夫转换 GARCH-M 模型．所示数字是基于 5 000+2 000 次迭代的吉布斯抽样的参数后验均值和标准差．前 5 000 次迭代的结果舍去．先验分布和初始参数估计值在文中给出．

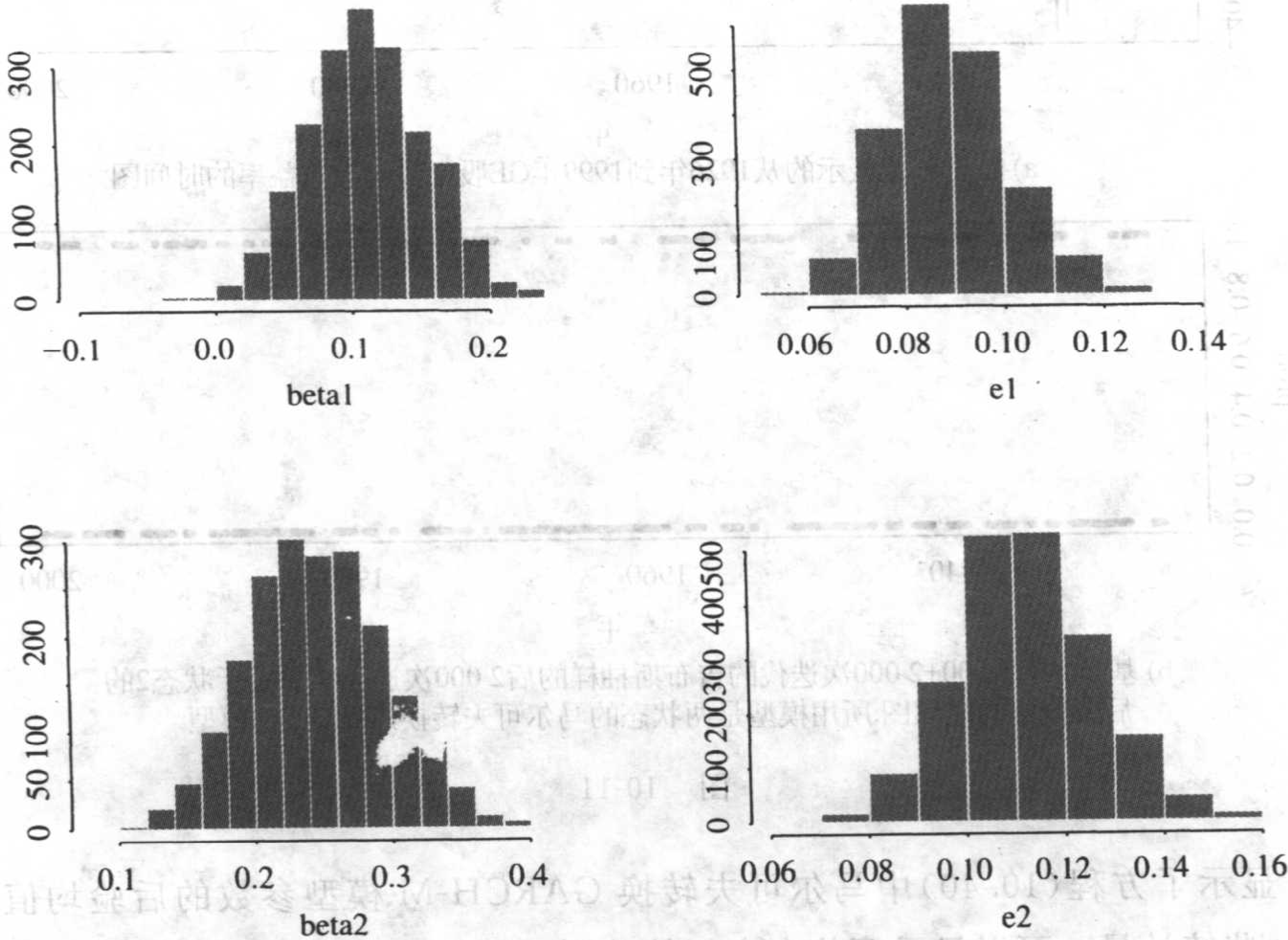


图 10-12　对 1926 年到 1999 年 GE 股票的月对数收益率拟合的两状态马尔可夫转换 GARCH-M 模型的转移概率和风险溢价的直方图．结果基于共有 5 000+2 000 次迭代的吉布斯抽样的后 2 000 次迭代

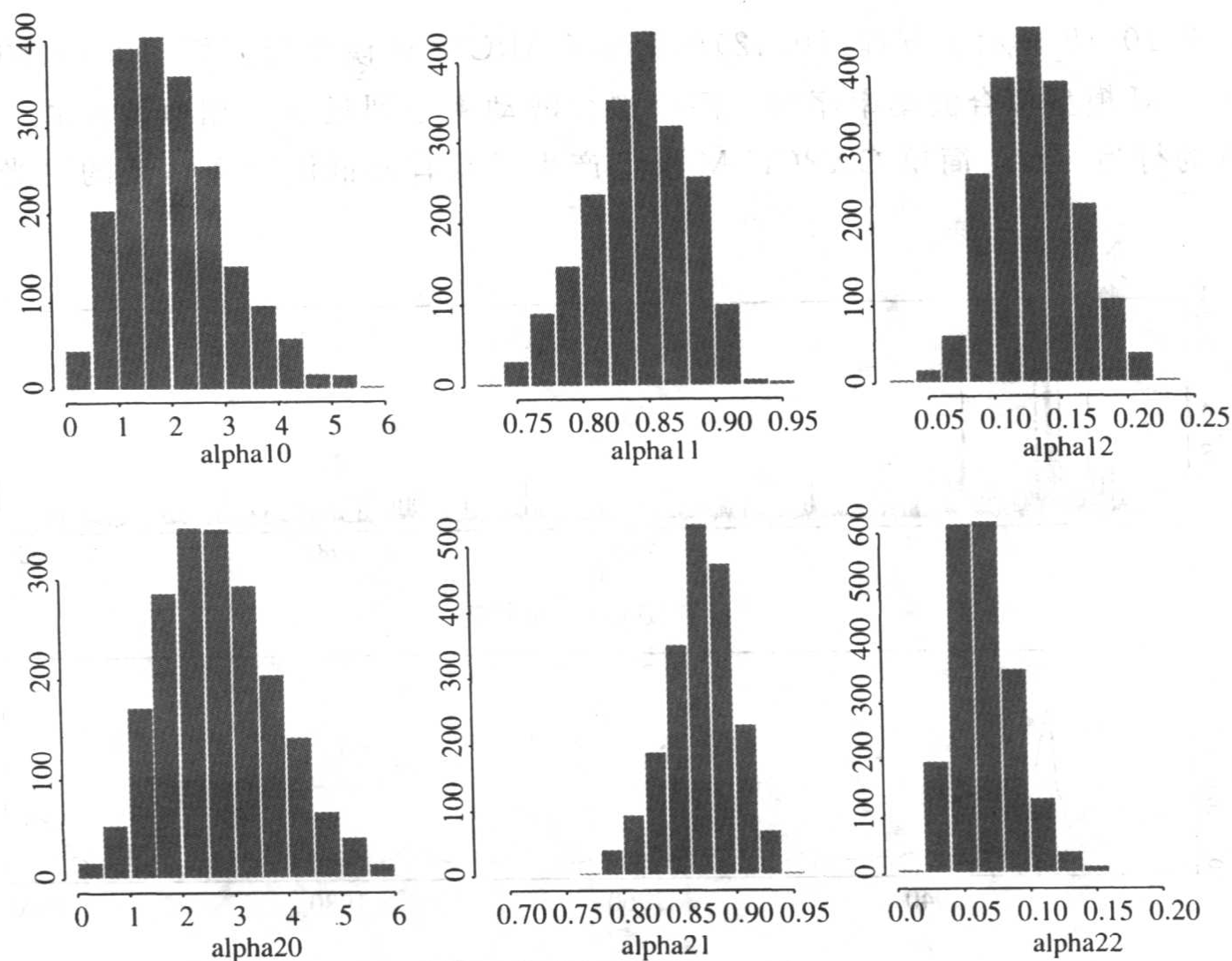


图 10-13 对 1926 年到 1999 年 GE 股票的月对数收益率拟合的两状态马尔可夫转换 GARCH-M 模型的波动率参数估计的直方图．结果基于共有 5 000+2 000 次迭代的吉布斯抽样的后 2 000 次迭代

436 ~ 437

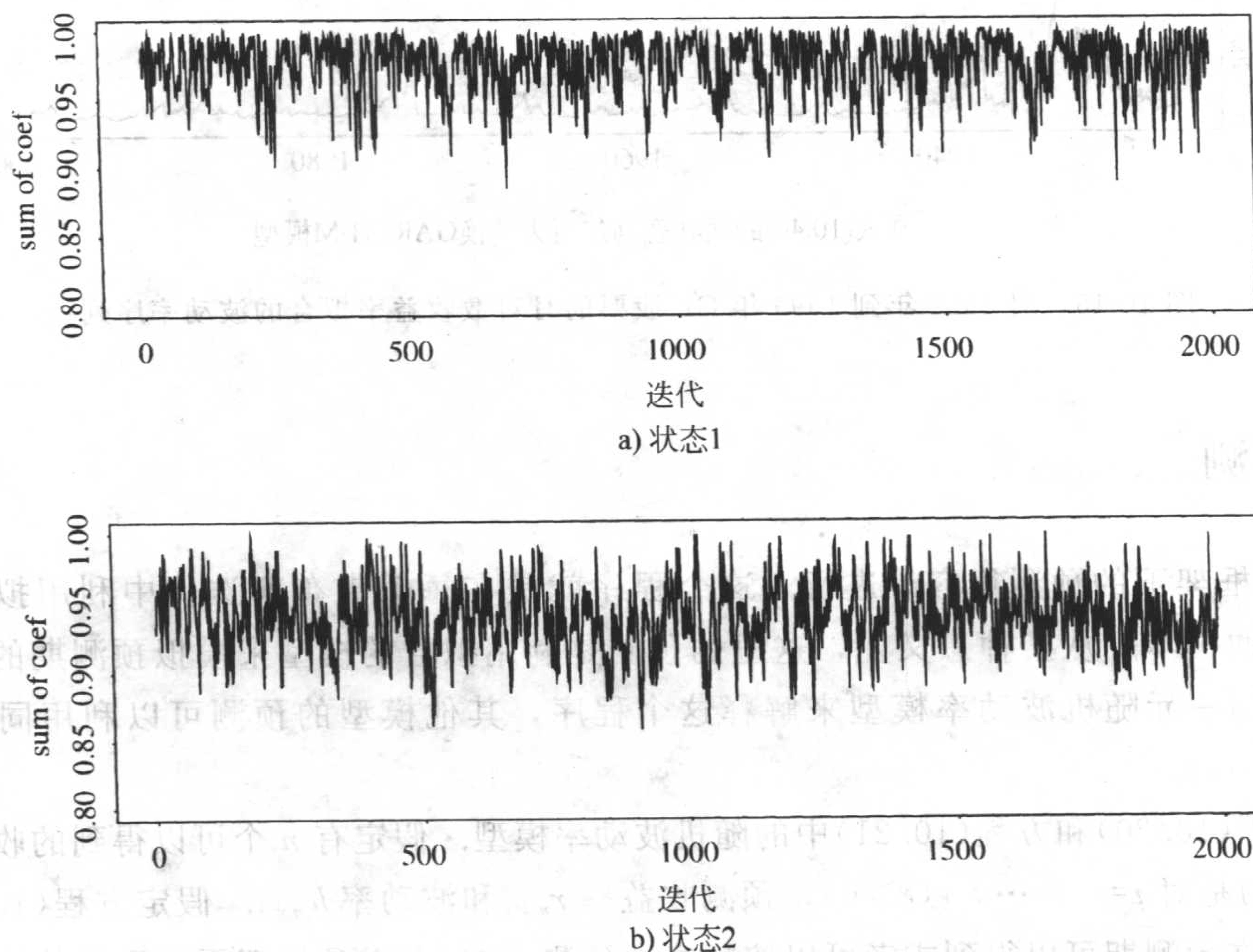


图 10-14 对 1926 年到 1999 年 GE 股票的月对数收益率拟合的两状态马尔可夫转换 GARCH-M 模型的持续参数 $\alpha_{i1}+\alpha_{i2}$ 的时间图．结果基于共有 5 000+2 000 次迭代的吉布斯抽样的后 2 000 次迭代

最后，图 10-15 显示了方程(10.42)的简单 GARCH-M 模型与方程(10.40)的马尔可夫转换 GARCH-M 模型拟合波动率序列. 两个拟合波动率序列显示了相似的模式，且与对数收益率平方的行为一致. 简单 GARCH-M 模型产生了具有较低波动率估计的更光滑的波动率序列.

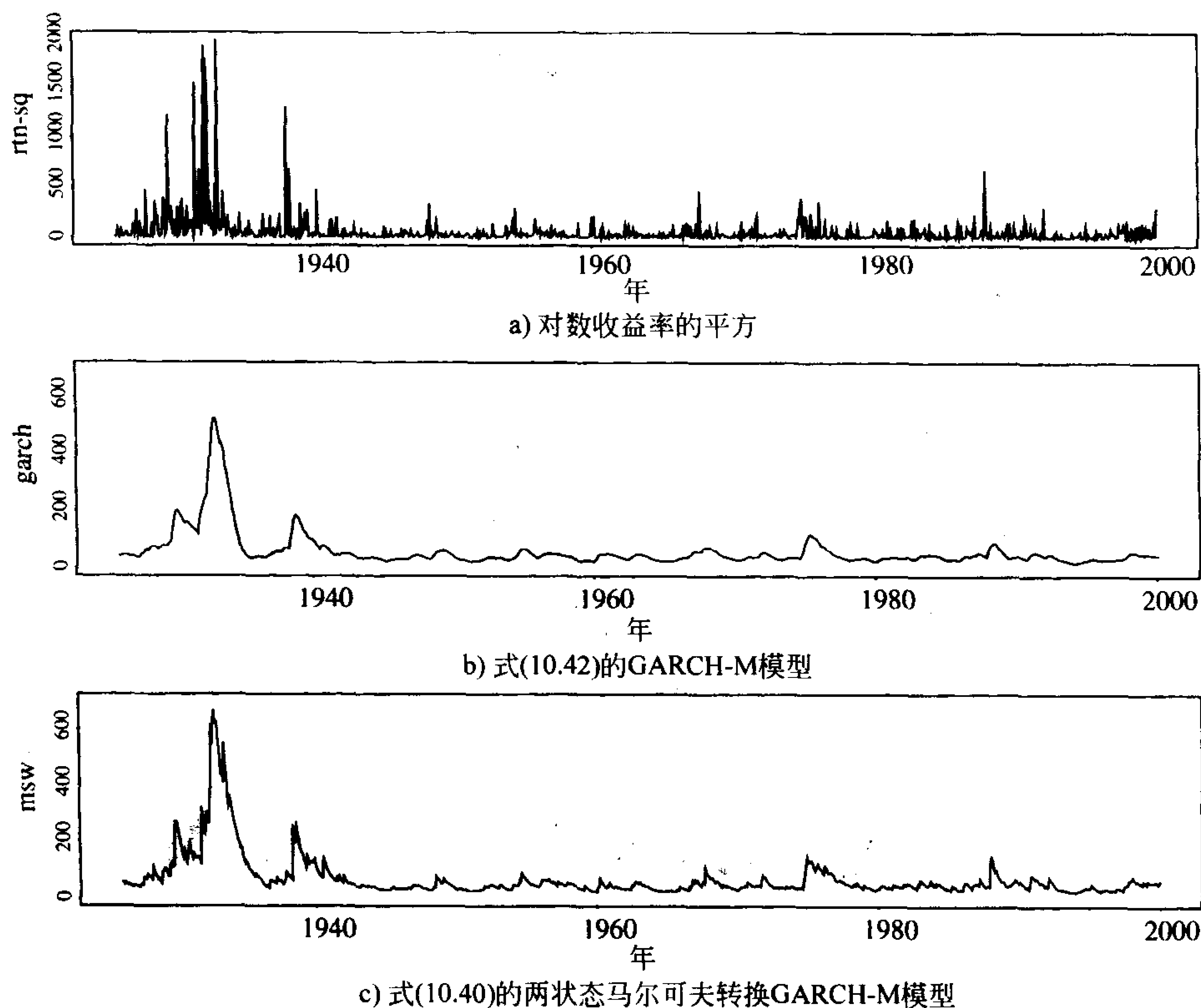


图 10-15 对 1926 年到 1999 年 GE 股票的月对数收益率拟合的波动率序列

10.9 预测

MCMC 框架下的预测很容易进行. 这个程序就是在每个吉布斯迭代中利用拟合的模型产生预测期的样本. 从某种意义上，这里的预测是利用拟合的模型来模拟预测期的实现进行的. 我们利用一元随机波动率模型来解释这个程序，其他模型的预测可以利用同样的方法得到.

考虑方程(10.20)和方程(10.21)中的随机波动率模型. 假定有 n 个可以得到的收益率，且我们感兴趣的是对 $i=1, \cdots, \ell(\ell>0)$，预测收益率 r_{n+i} 和波动率 h_{n+i}. 假定方程(10.20)中的解释变量 x_{jt} 在预测期可以得到或者可以按顺序地预测. 在 MCMC 框架下模型的估计是通过吉布斯抽样进行的，从它们的条件后验分布中迭代地抽取参数值. 将第 j 次吉布斯迭代中的参数

表示为$\boldsymbol{\beta}_j=(\beta_{0,j},\ \cdots,\ \beta_{p,j})'$，$\boldsymbol{\alpha}_j=(\alpha_{0,j},\ \alpha_{1,j})'$，和$\sigma^2_{v,j}$. 换句话说，在第$j$次吉布斯迭代中，模型为

$$r_t=\beta_{0,j}+\beta_{1,j}x_{1t}+\cdots+\beta_{p,j}x_{pt}+a_t \tag{10.43}$$

$$\ln h_t=\alpha_{0,j}+\alpha_{1,j}\ln h_{t-1}+v_t,\quad \mathrm{Var}(v_t)=\sigma^2_{v,j} \tag{10.44}$$

438~439

我们可以利用这个模型来产生r_{n+i}和h_{n+i}($i=1,\ \cdots,\ l$)的一个实现，并将模拟的实现分别表示为$r_{n+i,j}$和$h_{n+i,j}$. 这些实现如下产生：

- 从$N(0,\ \sigma^2_{v,j})$中抽取一个随机样本v_{n+1}，利用方程(10.44)计算$h_{n+i,j}$.
- 从$N(0,\ 1)$中抽取一个随机样本ε_{n+1}，得到$a_{n+1,j}=\sqrt{h_{n+1,j}}\varepsilon_{n+1}$，并利用方程(10.43)计算$r_{n+1,j}$.
- 对$n+i$，$i=2,\ \cdots,\ \ell$，按顺序重复前面的两个步骤.

如果在模型估计中运行$M+N$个迭代的吉布斯抽样，我们仅仅需要计算最后N个迭代的预测. 这就得到r_{n+i}和h_{n+i}的一个随机样本. 具体地，我们得到

$$\{r_{n+1,j},\cdots,r_{n+\ell,j}\}_{j=1}^{N},\quad \{h_{n+1,j},\cdots,h_{n+\ell,j}\}_{j=1}^{N}.$$

可以利用这两个随机样本进行推断. 例如，收益率r_{n+i}和波动率h_{n+i}的点预测就是这两个随机样本的样本均值. 类似地，可以利用样本标准差作为预测误差的方差. 为了改善波动率预测中的计算有效性，可以利用重要抽样；见 Gelman，Carlin，Stern 和 Rubin(1995，第307页).

例 10.6(例 10.3 的续) 作为说明，我们考虑标准普尔 500 指数从 1962 年至 1999 年的月对数收益率序列. 表 10-4 给出了从 1999 年 12 月开始的五个预测期的收益率及其波动率的点预测. 方程(10.26)中的 GARCH 模型与方程(10.27)中的随机波动率模型都用来预测. GARCH(1，1)模型的波动率预测随着预测时间区间的加大而渐近地增加至无条件方差 3.349/(1－0.086－0.735)＝18.78. 用随机波动率模型得到的波动率预测值比用 GARCH 模型得到的预测值高. 这是可以理解的，因为随机波动率模型产生预测时考虑了参数的不确定性. 在比较中，GARCH 模型假定参数固定，且在方程(10.26)中给出. 这是一个重要的差别，也是 GARCH 模型与衍生产品定价得到的隐含波动率相比可能低估波动率的一个原因.

表 10-4

(a)对数收益率					
预测步长	1	2	3	4	5
GARCH	0.66	0.66	0.66	0.66	0.66
SVM	0.53	0.78	0.92	0.88	0.84
(b)波动率					
预测步长	1	2	3	4	5
GARCH	17.98	18.12	18.24	18.34	18.42
SVM	19.31	19.36	19.35	19.65	20.13

注：标准普尔 500 指数的月对数收益率的波动率预测. 数据的时间段是 1962 年 1 月至 1999 年 12 月，预测原点是 1999 年 12 月. 随机波动率模型的预测是由 2 000+2 000 次迭代的吉布斯抽样得到的.

注释：除了在预测中考虑了参数不确定性的优势，MCMC 方法还有效地产生了所研究的波动率的一个可预测分布. 预测分布比简单的点预测带有更多信息. 例如，它可以用来得到 VaR 计算中所需要的分位数.

10.10 其他应用

MCMC 方法对许多其他的金融问题也是适用的. 例如，Zhang，Russell 和 Tsay(2000)用它来分析买卖报价中的信息决定性，McCulloch 和 Tsay(2000)用此方法来对 IBM 交易数据估计一个等级模型. Eraker(2001)，Elerian，Chib 和 Shephard(2001)用它来估计扩散方程. 这个方法在 VaR 计算中也是有用的，因为它提供了估计预测分布的一个自然方法. 主要的问题不是这个方法是否在大多数金融应用中都可以利用，而是方法可以变得多么有效. 只有时间和经验可以提供该问题一个充分答案.

练习题

1. 假设 x 服从均值为 μ、方差为 4 的正态分布. 假定 μ 的先验分布也是均值为 0、方差为 25 的正态分布. 给定数据点 x，μ 的后验分布是什么？
2. 考虑 10.5 节中带时间序列误差的线性回归模型. 假定 z_t 是一个 AR(p)过程(即 $z_t=\phi_1 z_{t-1}+\cdots+\phi_p z_{t-p}+a_t$). 令 $\boldsymbol{\phi}=(\phi_1,\cdots,\phi_p)'$ 表示 AR 参数向量. 假设共轭先验分布为 $\boldsymbol{\beta}\sim N(\boldsymbol{\beta}_o,\Sigma_o)$，$\boldsymbol{\phi}\sim N(\boldsymbol{\phi}_o,\boldsymbol{A}_o)$，$(v\lambda)/\sigma^2-\chi^2_v$，导出条件后验分布 $f(\boldsymbol{\beta}\mid \boldsymbol{Y},\boldsymbol{X},\boldsymbol{\phi},\sigma^2)$，$f(\boldsymbol{\phi}\mid \boldsymbol{Y},\boldsymbol{X},\boldsymbol{\beta},\sigma^2)$ 和 $f(\sigma^2\mid \boldsymbol{Y},\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{\phi})$.

440 ～ 441

3. 考虑 10.6.1 节中的线性 AR(p)模型. 假定 x_h 和 x_{h+1} 是两个缺失值，它们的联合先验分布是均值为 $\boldsymbol{\mu}_0$、协方差阵为 Σ_0 的多元正态分布. 其他先验分布如本章中所述. 两个缺失值的条件后验分布是什么？
4. 考虑 General Motors 股票从 1950 年到 1999 年的月对数收益率的 600 个观测值.
   (a)对此序列建立一个 GARCH 模型；
   (b)对此序列建立一个随机波动率模型；
   (c)比较两个波动率模型.
5. 对 Cisco Systems 股票从 1991 年 1 月到 1999 年 12 月的日对数收益率建立一个随机波动率模型. 你可以从 CRSP 数据库或文件“d-csco9199.dat”中下载数据. 利用所建的模型得到以 1999 年 12 月为预测原点的向前 1 步的波动率预测分布. 最后，用所得的预测分布计算下一交易日的 100 万美元多头头寸的概率为 0.01 的 VaR 值.
6. 对 General Motors 股票和标准普尔 500 指数的月对数收益率建立一个二元随机波动率模型，样本时间段为 1950 年 1 月到 1999 年 12 月. 讨论两个波动率过程的关系，并计算 GM 股票的时变贝塔.

参考文献

Box, G. E. P., and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis.* Addison-Wesley: Reading, MA.

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of time series parameters in the presence of outliers," *Technometrics*, **30**, 193–204.

Carlin, B. P., and Louis, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed., Chapman and Hall: London.

DeGroot, M. H. (1970), *Optimal Statistical Decisions*, McGraw-Hill: New York.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm" (with discussion), *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Elerian, O., Chib, S., and Shephard, N. (2001), "Likelihood inference for discretely observed nonlinear diffusions," *Econometrica*, **69**, 959–993.

Eraker, B. (2001), "Markov Chain Monte Carlo analysis of diffusion models with application to finance," *Journal of Business & Economic Statistics* **19**, 177–191.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, **85**, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, CRC Press: London.

Geman, S., and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Hasting, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, **57**, 97–109.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian analysis of stochastic volatility models" (with discussion), *Journal of Business & Economic Statistics*, **12**, 371–417.

Jones, R. H.(1980), "Maximum likelihood fitting of ARMA models to time series with missing observations," *Technometrics*, **22**, 389–395.

442

Justel, A., Peña, D., and Tsay, R. S. (2001), "Detection of outlier patches in autoregressive time series," *Statistica Sinica*, **11** (to appear).

Liu, J., Wong, W. H., and Kong, A. (1994), "Correlation structure and convergence rate of the Gibbs samplers I: Applications to the comparison of estimators and augmentation schemes," *Biometrika*, **81**, 27–40.

McCulloch, R. E., and Tsay, R. S. (1994), "Bayesian analysis of autoregressive time series via the Gibbs sampler," *Journal of Time Series Analysis*, **15**, 235–250.

McCulloch, R. E., and Tsay, R. S. (1994), "Statistical analysis of economic time series via Markov switching models," *Journal of Time Series Analysis*, **15**, 523–539.

Metropolis, N., and Ulam, S. (1949), "The Monte Carlo method," *Journal of the American Statistical Association*, **44**, 335–341.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, **21**, 1087–1092.

Tanner, M. A. (1996), *Tools for Statistical Inference: Methods for the exploration of posterior distributions and likelihood functions*, 3rd ed., Springer-Verlag: New York.

Tanner, M. A., and Wong, W. H. (1987), "The calculation of posterior distributions by data augmentation" (with discussion), *Journal of the American Statistical Association*, **82**, 528–550.

Tierney, L. (1994), "Markov chains for exploring posterior distributions" (with discussion), *Annals of Statistics*, **22**, 1701–1762.

Tsay, R. S. (1988), "Outliers, level shifts, and variance changes in time series," *Journal of Forecasting*, **7**, 1–20.

Tsay, R. S., Peña, D., and Pankratz, A. (2000), "Outliers in multivariate time series," *Biometrika*, **87**, 789–804.

Zhang, M. Y., Russell, J. R. and Tsay, R. S. (2000), "Determinants of bid and ask quotes and implications for the cost of trading," Working paper, Statistics Research Center, Graduate School of Business, University of Chicago.

443

索　引

索引中的页码为英文原书页码，与书中页边标注的页码一致.